1368.61P

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04493 A2

(51) International Patent Classification[7]: C07K 14/155

(21) International Application Number: PCT/US01/21241

(22) International Filing Date: 5 July 2001 (05.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/610,313 5 July 2000 (05.07.2000) US

(71) Applicants *(for all designated States except US)*: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US). **UNIVERSITY OF STELLENBOSCH** [ZA/ZA]; P.O. Box 19063, 7505 Tygerberg (ZA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ZUR MEGEDE, Jan** [DE/US]; c/o Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). **BARNETT, Susan, W.** [US/US]; c/o Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). **ENGELBRECHT, Susan** [ZA/ZA]; c/o University of Stellenbosch, P.O. Box 19063, 7505 Tygerberg (ZA). **VAN RENSBURG, Estrelita, Janse** [ZA/ZA]; c/o University of Stellenbosch, P.O. Box 19063, 7505 Tygerberg (ZA).

(74) Agents: **DOLLARD, Anne, S.** et al.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/04493 A2

(54) Title: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

(57) Abstract: The present invention relates to polynucleotides encoding immunogenic HIV type C polypeptides. Uses of the polynucleotides in applications including DNA immunization, generation of packaging cell lines, and production of HIV Type C proteins are also described.

POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

TECHNICAL FIELD

Polynucleotides encoding antigenic Type C HIV polypeptides (*e.g.*, Gag, pol, vif, vpr, tat, rev, vpu, env, and nef) are described, as are uses of these polynucleotides and polypeptide products in immunogenic compositions. Also described are polynucleotide sequences from South African variants of HIV Type C.

BACKGROUND OF THE INVENTION

Acquired immune deficiency syndrome (AIDS) is recognized as one of the greatest health threats facing modern medicine. There is, as yet, no cure for this disease. In 1983-1984, three groups independently identified the suspected etiological agent of AIDS. See, e.g., Barre-Sinoussi et al. (1983) Science 220:868-871; Montagnier et al., in Human T-Cell Leukemia Viruses (Gallo, Essex & Gross, eds., 1984); Vilmer et al. (1984) The Lancet 1:753; Popovic et al. (1984) Science 224:497-500; Levy et al. (1984) Science 225:840-842. These isolates were variously called lymphadenopathy-associated virus (LAV), human T-cell lymphotropic virus type III (HTLV-III), or AIDS-associated retrovirus (ARV). All of these isolates are strains of the same virus, and were later collectively named Human Immunodeficiency-Virus (HIV). With the isolation of a related AIDS-causing virus, the strains originally called HIV are now termed HIV-1 and the related virus is called HIV-2 See, e.g., Guyader et al. (1987) Nature 326:662-669; Brun-Vezinet et al. (1986) Science 233:343-346; Clavel et al. (1986) Nature 324:691-695.

A great deal of information has been gathered about the HIV virus, however, to date an effective vaccine has not been identified. Several targets for vaccine development have been examined including the *env* and *Gag* gene products encoded by HIV. Gag gene products include, but are not limited to, Gag-polymerase and Gag-protease. Env gene products include, but are not limited to, monomeric gp120 polypeptides, oligomeric gp140 polypeptides and gp160 polypeptides.

Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.* 72(2):1497-1503, 1998) described an increased

immune response elicited by DNA vaccination employing a synthetic gp120 sequence with modified codon usage. Schneider, et al., (*J Virol.* 71(7):4892-4903, 1997) discuss inactivation of inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences.

The *Gag* proteins of HIV-1 are necessary for the assembly of virus-like particles. HIV-1 *Gag* proteins are involved in many stages of the life cycle of the virus including, assembly, virion maturation after particle release, and early post-entry steps in virus replication. The roles of HIV-1 *Gag* proteins are numerous and complex (Freed, E.O., *Virology* **251**:1-15, 1998).

Wolf, et al., (PCT International Application, WO 96/30523, published 3 October 1996; European Patent Application, Publication No. 0 449 116 A1, published 2 October 1991) have described the use of altered pr55 *Gag* of HIV-1 to act as a non-infectious retroviral-like particulate carrier, in particular, for the presentation of immunologically important epitopes. Wang, et al., (*Virology* **200**:524-534, 1994) describe a system to study assembly of HIV Gag-β-galactosidase fusion proteins into virions. They describe the construction of sequences encoding HIV Gag-β-galactosidase fusion proteins, the expression of such sequences in the presence of HIV Gag proteins, and assembly of these proteins into virus particles.

Shiver, et al., (PCT International Application, WO 98/34640, published 13 August 1998) described altering HIV-1 (CAM1) *Gag* coding sequences to produce synthetic DNA molecules encoding HIV *Gag* and modifications of HIV *Gag*. The codons of the synthetic molecules were codons preferred by a projected host cell.

Recently, use of HIV Env polypeptides in immunogenic compositions has been described. (see, U.S. Patent No. 5,846,546 to Hurwitz et al., issued December 8, 1998, describing immunogenic compositions comprising a mixture of at least four different recombinant virus that each express a different HIV env variant; and U.S. Patent No. 5,840,313 to Vahlne et al., issued November 24, 1998, describing peptides which correspond to epitopes of the HIV-1 gp120 protein). In addition, U.S. Patent No. 5,876,731 to Sia et al, issued March 2, 1999 describes candidate vaccines against HIV comprising an amino acid sequence of a T-cell epitope of Gag linked directly to an amino acid sequence of a B-cell epitope of the V3 loop protein of an HIV-1 isolate containing the sequence GPGR. There remains a need for antigenic HIV polypeptides, particularly Type C isolates.

SUMMARY OF THE INVENTION

Described herein are novel Type C HIV sequences, for example, 8_5_TV1_C.ZA, 8_2_TV1_C.ZA and 12-5_1_TV2_C.ZA, polypeptides encoded by these novel sequences, and synthetic expression cassettes generated from these and other Type C HIV sequences.

In certain embodiments, the present invention relates synthetic expression cassettes encoding HIV Type C polypeptides, including Env, Gag, Pol, Prot, Vpr, Vpu, Vif, Nef, Tat, Rev and/or fragments thereof. In addition, the present invention also relates to improved expression of HIV Type C polypeptides and production of virus-like particles. Synthetic expression cassettes encoding the HIV polypeptides (*e.g.*, Gag-, pol-, protease (prot)-, reverse transcriptase, integrase, RNAseH, Tat, Rev, Nef, Vpr, Vpu, Vif and/or Env- containing polypeptides) are described, as are uses of the expression cassettes.

Thus, one aspect of the present invention relates to expression cassettes and polynucleotides contained therein. The expression cassettes typically include an HIV-polypeptide encoding sequence inserted into an expression vector backbone. In one embodiment, an expression cassette comprises a polynucleotide sequence encoding one or more *Pol*-containing polypeptides, wherein the polynucleotide sequence comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and more preferably about 98% sequence (and any integers between these values) identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Pol*-containing polypeptides include, but are not limited to, those shown in SEQ ID NO:30, SEQ ID NO:31; SEQ ID NO:32; SEQ ID NO:62; SEQ ID NO:103; SEQ ID NO:58; SEQ ID NO:60; SEQ ID NO:64; SEQ ID NO:66; SEQ ID NO:68; SEQ ID NO:70; SEQ ID NO:76; and SEQ ID NO:78.

The polynucleotides encoding the HIV polypeptides of the present invention may also include sequences encoding additional polypeptides. Such additional polynucleotides encoding polypeptides may include, for example, coding sequences for other viral proteins (*e.g.*, hepatitis B or C or other HIV proteins, such as, polynucleotide sequences encoding an HIV *Gag* polypeptide, polynucleotide sequences encoding an HIV *Env* polypeptide and/or polynucleotides encoding one or more of vif, vpr, tat, rev, vpu and nef); cytokines or other transgenes. In one embodiment, the sequence encoding the HIV *Pol* polypeptide(s) can be modified by deletions of coding regions corresponding to reverse transcriptase and integrase.

Such deletions in the polymerase polypeptide can also be made such that the polynucleotide sequence preserves T-helper cell and CTL epitopes. Other antigens of interest may be inserted into the polymerase as well.

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Gag*-containing polypeptide, wherein the polynucleotide sequence encoding the *Gag* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Gag*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 844-903 of Figure 1 (a Gag major homology region) (SEQ ID NO:1); nucleotides 841-900 of Figure 2 (a Gag major homology region) (SEQ ID NO:2); Figure 24 (SEQ ID NO:53, a Gag major homology region); the sequence presented as Figure 1 (SEQ ID NO:3); the sequence presented as Figure 22 (SEQ ID NO:51); the sequence presented as Figure 70 (SEQ ID NO:99); and the sequence presented as Figure 2 (SEQ ID NO:4). As noted above, the polynucleotides encoding the *Gag*-containing polypeptides of the present invention may also include sequences encoding additional polypeptides.

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Env*-containing polypeptide, wherein the polynucleotide sequence encoding the *Env* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Env*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 1213-1353 of Figure 3 (SEQ ID NO:5) (encoding an Env common region); the sequence presented as Figure 17 (SEQ ID NO:46) (encoding a 97 nucleotide long Env common region); SEQ ID NO:47 (encoding a 144 nucleotide long Env common region); nucleotides 82-1512 of Figure 3 (SEQ ID NO:6) (encoding a gp120 polypeptide); nucleotides 82-2025 of Figure 3 (SEQ ID NO:7) (encoding a gp140 polypeptide); nucleotides 82-2547 of Figure 3 (SEQ ID NO:8) (encoding a gp160 polypeptide); SEQ ID NO:49 (encoding a gp160 polypeptide); nucleotides 1-2547 of Figure 3 (SEQ ID NO:9) (encoding a gp160 polypeptide with signal sequence); nucleotides 1513-2547 of Figure 3 (SEQ ID NO:10) (encoding a gp41 polypeptide); nucleotides 1210-1353 of

Figure 4 (SEQ ID NO:11) (encoding an Env common region); nucleotides 73-1509 of Figure 4 (SEQ ID NO:12) (encoding a gp120 polypeptide); nucleotides 73-2022 of Figure 4 (SEQ ID NO:13) (encoding a gp140 polypeptide); nucleotides 73-2565 of Figure 4 (SEQ ID NO:14) (encoding a gp160 polypeptide); nucleotides 1-2565 of Figure 4 (SEQ ID NO:15) (encoding a gp160 polypeptide with signal sequence); the sequence presented as Figure 20 (SEQ ID NO:49) (encoding a gp160 polypeptide); the sequence presented as Figure 68 (SEQ ID NO:97) (encoding a gp160 polypeptide); nucleotides 1510-2565 of Figure 4 (SEQ ID NO:16) (encoding a gp41 polypeptide); nucleotides 7 to 1464 of Figure 90 (SEQ ID NO:119) (encoding a gp120 polypeptide with modified wild type signal sequence); nucleotides 7 to 1977 of Figure 91 (SEQ ID NO:120) (encoding a gp140 polypeptide including signal sequence modified from wild-type 8_2_TV1_C.ZA (*e.g.*, "modified wild type leader sequence")); nucleotides 7 to 1977 of Figure 92 (SEQ ID NO:121) (encoding a gp140 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2388 of Figure 93 (SEQ ID NO:122) (encoding a gp160 polypeptide with modified wild type signal sequence); nucleotides 7 to 2520 of Figure 94 (SEQ ID NO:123) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2520 of Figure 95 (SEQ ID NO:124) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 13 to 2604 of Figure 96 (SEQ ID NO:125) (encoding a gp160 polypeptide with TPA1 signal sequence); nucleotides 7 to 2607 of Figure 97 (SEQ ID NO:126) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 1 to 2049 of Figure 100 (SEQ ID NO:131) (encoding a gp140 polypeptide with TPA1 signal sequence); nucleotides 7 to 1607 of Figure 98 (SEQ ID NO:126) (encoding a gp160 polypeptide with wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2064 of SEQ ID NO:132 (encoding a gp140 polypeptide with modified wild-type 8_2_TV1_C.ZA leader sequence); and nucleotides 7 to 2064 of SEQ ID NO:133 (encoding a gp140 polypeptide with wild-type 8_2_TV1_C.ZA leader sequence).

In certain embodiments, the Env-encoding sequences will contain further modifications, for instance mutation of the cleavage site to prevent the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide (SEQ ID NO:121 and SEQ ID NO:124) or deletion of variable regions V1 and/or V2 (SEQ ID NO:119; SEQ ID NO:120; SEQ ID NO:121; SEQ ID NO:122; SEQ ID NO:123; and SEQ ID NO:124).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Nef*-containing polypeptide, wherein the polynucleotide sequence encoding the *Nef* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Nef*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 26 (SEQ ID NO:55); the sequence presented in Figure 72 (SEQ ID NO:101); the sequence presented in Figure 28 (SEQ ID NO:57); the sequence presented in Figure 67 (SEQ ID NO:96); the sequence presented in Figure 103 (SEQ ID NO:134); and the sequence presented in Figure 104 (SEQ ID NO:135).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Rev*-containing polypeptide, wherein the polynucleotide sequence encoding the *Rev* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Rev*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 43 (SEQ ID NO:72); the sequence presented in Figure 76 (SEQ ID NO:105); the sequence presented in Figure 45 (SEQ ID NO:74); the sequence presented in Figure 78 (SEQ ID NO:107); and the sequence presented in Figure 62 (SEQ ID NO:91).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Tat*-containing polypeptide, wherein the polynucleotide sequence encoding the *Tat* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Tat*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 51 (SEQ ID NO:80); the sequence presented in Figure 80 (SEQ ID NO:109); the sequence presented in Figure 52 (SEQ ID NO:81); the sequence presented in Figure 54 (SEQ ID NO:83); and the sequence presented in Figure 82 (SEQ ID NO:111).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Vif*-containing polypeptide, wherein the polynucleotide sequence encoding the *Vif* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Vif*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 56 (SEQ ID NO:85); and the sequence presented in Figure 84 (SEQ ID NO:113).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Vpr*-containing polypeptide, wherein the polynucleotide sequence encoding the *Vpr* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Vpr*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 58 (SEQ ID NO:87); and the sequence presented in Figure 86 (SEQ ID NO:115).

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Vpu*-containing polypeptide, wherein the polynucleotide sequence encoding the *Vpu* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Vpu*-containing polypeptides include, but are not limited to, the following polynucleotides: the sequence presented in Figure 60 (SEQ ID NO:89); and the sequence presented in Figure 88 (SEQ ID NO:117).

Further embodiments of the present invention include purified polynucleotides of any of the sequences described herein. Exemplary polynucleotide sequences encoding *Gag*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 844-903 of Figure 1 (SEQ ID NO:1) (a Gag major homology region); nucleotides 841-900 of Figure 2 (SEQ ID NO:2) (a Gag major homology region); the sequence presented as Figure 1 (SEQ ID NO:3); the sequence presented as Figure 2 (SEQ ID NO:4); the sequence presented as Figure 22 (SEQ ID NO:51); the sequence presented as Figure 70 (SEQ

ID NO:99); and the sequence presented as Figure 24 (SEQ ID NO:53) (a Gag major homology region).

Exemplary polynucleotide sequences encoding *Env*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 1213-1353 of Figure 3 (SEQ ID NO:5) (encoding an Env common region); the sequence presented as Figure 17 (SEQ ID NO:46) (encoding a 97 nucleotide long Env common region); SEQ ID NO:47 (encoding a 144 nucleotide long Env common region); nucleotides 82-1512 of Figure 3 (SEQ ID NO:6) (encoding a gp120 polypeptide); nucleotides 82-2025 of Figure 3 (SEQ ID NO:7) (encoding a gp140 polypeptide); nucleotides 82-2547 of Figure 3 (SEQ ID NO:8) (encoding a gp160 polypeptide); SEQ ID NO:49 (encoding a gp160 polypeptide); nucleotides 1-2547 of Figure 3 (SEQ ID NO:9) (encoding a gp160 polypeptide with signal sequence); nucleotides 1513-2547 of Figure 3 (SEQ ID NO:10) (encoding a gp41 polypeptide); nucleotides 1210-1353 of Figure 4 (SEQ ID NO:11) (encoding an Env common region); nucleotides 73-1509 of Figure 4 (SEQ ID NO:12) (encoding a gp120 polypeptide); nucleotides 73-2022 of Figure 4 (SEQ ID NO:13) (encoding a gp140 polypeptide); nucleotides 73-2565 of Figure 4 (SEQ ID NO:14) (encoding a gp160 polypeptide); nucleotides 1-2565 of Figure 4 (SEQ ID NO:15) (encoding a gp160 polypeptide with signal sequence); the sequence presented as Figure 20 (SEQ ID NO:49) (encoding a gp160 polypeptide); the sequence presented as Figure 68 (SEQ ID NO:97) (encoding a gp160 polypeptide); nucleotides 1510-2565 of Figure 4 (SEQ ID NO:16) (encoding a gp41 polypeptide); nucleotides 7 to 1464 of Figure 90 (SEQ ID NO:119) (encoding a gp120 polypeptide with modified wild type signal sequence); nucleotides 7 to 1977 of Figure 91 (SEQ ID NO:120) (encoding a gp140 polypeptide including signal sequence modified from wild-type 8_2_TV1_C.ZA (*e.g.*, "modified wild type leader sequence")); nucleotides 7 to 1977 of Figure 92 (SEQ ID NO:121) (encoding a gp140 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2388 of Figure 93 (SEQ ID NO:122) (encoding a gp160 polypeptide with modified wild type signal sequence); nucleotides 7 to 2520 of Figure 94 (SEQ ID NO:123) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2520 of Figure 95 (SEQ ID NO:124) (encoding a gp160 polypeptide with modified wild type 8_2_TV1_C.ZA signal sequence); nucleotides 13 to 2604 of Figure 96 (SEQ ID NO:125) (encoding a gp160 polypeptide with TPA1 signal sequence); nucleotides 7 to 2607 of Figure 97 (SEQ ID NO:126) (encoding a gp160 polypeptide with modified wild type

8_2_TV1_C.ZA signal sequence); nucleotides 1 to 2049 of Figure 100 (SEQ ID NO:131) (encoding a gp140 polypeptide with TPA1 signal sequence); nucleotides 7 to 1607 of Figure 98 (SEQ ID NO:126) (encoding a gp160 polypeptide with wild type 8_2_TV1_C.ZA signal sequence); nucleotides 7 to 2064 of SEQ ID NO:132 (encoding a gp140 polypeptide with modified wild-type 8_2_TV1_C.ZA leader sequence); and nucleotides 7 to 2064 of SEQ ID NO:133 (encoding a gp140 polypeptide with wild-type 8_2_TV1_C.ZA leader sequence).

Exemplary purified polynucleotides encoding additional HIV polynucleotides include: Pol-encoding polynucleotides (*e.g.*, SEQ ID NO:30, SEQ ID NO:31; SEQ ID NO:32; SEQ ID NO:62; SEQ ID NO:103; SEQ ID NO:58; SEQ ID NO:60; SEQ ID NO:64; SEQ ID NO:66; SEQ ID NO:68; SEQ ID NO:70; SEQ ID NO:76; and SEQ ID NO:78); Nef-encoding polynucleotides (*e.g.*, SEQ ID NO:55; SEQ ID NO:101; SEQ ID NO:57; SEQ ID NO:96); Rev-encoding polynucleotides (*e.g.*, SEQ ID NO:72; SEQ ID NO:105; SEQ ID NO:74); SEQ ID NO:107; SEQ ID NO:91); Tat-encoding polynucleotides (*e.g.*, SEQ ID NO:80; SEQ ID NO:109; SEQ ID NO:81; SEQ ID NO:83; SEQ ID NO:111); Vif-encoding polynucleotides (*e.g.*,SEQ ID NO:85; SEQ ID NO:113); and Vpr-encoding polynucleotides (*e.g.*,SEQ ID NO:87; SEQ ID NO:115); Vpu-encoding polynucleotides (*e.g.*,SEQ ID NO:89; SEQ ID NO:117).

In other embodiments, the present invention relates to native HIV polypeptide-encoding sequences obtained from novel Type C strains; fragments of these native sequences; expression cassettes containing these wild-type sequences; and uses of these sequences, fragments and expression cassettes. Exemplary full length sequences are shown in SEQ ID NO:33 and SEQ ID NO:45. Exemplary fragments coding for various HIV gene products include: the sequence presented in Figure 19 (SEQ ID NO:48) (an Env-encoding sequence); the sequence presented in Figure 69 (SEQ ID NO:98) (an Env-encoding sequence); the sequence presented in Figure 21 (SEQ ID NO:50) (a gp160 polypeptide); the sequence presented in Figure 23 (SEQ ID NO:52) (a Gag polypeptide); the sequence presented in Figure 71 (SEQ ID NO:100) (a Gag polypeptide); the sequence presented in Figure 25 (SEQ ID NO:54) (a Gag polypeptide); the sequence presented in Figure 27 (SEQ ID NO:56) (a Nef polypeptide); the sequence presented in Figure 73 (SEQ ID NO:102) (a Nef polypeptide); the sequence presented in Figure 30 (SEQ ID NO:59) (a p15RNAseH polypeptide); the sequence presented in Figure 32 (SEQ ID NO:61) (a p31Integrase polypeptide); the sequence presented in Figure 34 (SEQ ID NO:63) (a Pol polypeptide); the sequence presented in Figure 75 (SEQ

ID NO:104) (a Pol polypeptide); the sequence presented in Figure 36 (SEQ ID NO:65) (a Prot polypeptide); the sequence presented in Figure 38 (SEQ ID NO:67) (a inactivated Prot polypeptide); the sequence presented in Figure 40 (SEQ ID NO:69) (an inactivated Prot and RT polypeptide); the sequence presented in Figure 42 (SEQ ID NO:71) (a Prot and RT polypeptide); the sequence presented in Figure 44 (SEQ ID NO:73) (a Rev polypeptide); the sequence presented in Figure 77 (SEQ ID NO:106) (a Rev polypeptide); the sequence presented in Figure 46 (SEQ ID NO:75) (a Rev polypeptide); the sequence presented in Figure 79 (SEQ ID NO:108) (a Rev polypeptide); the sequence presented in Figure 48 (SEQ ID NO:77) (an RT polypeptide); the sequence presented in Figure 50 (SEQ ID NO:79) (a mutated RT polypeptide); the sequence presented in Figure 53 (SEQ ID NO:82) (a Tat polypeptide); the sequence presented in Figure 81 (SEQ ID NO:110) (a Tat polypeptide); the sequence presented in Figure 55 (SEQ ID NO:84) (a Tat polypeptide); the sequence presented in Figure 83 (SEQ ID NO:112) (a Tat polypeptide); the sequence presented in Figure 57 (SEQ ID NO:86) (a Vif polypeptide); the sequence presented in Figure 85 (SEQ ID NO:114) (a Vif polypeptide); the sequence presented in Figure 59 (SEQ ID NO:88) (a Vpr polypeptide); the sequence presented in Figure 82 (SEQ ID NO:116) (a Vpr polypeptide); the sequence presented in Figure 61 (SEQ ID NO:90) (a Vpu polypeptide); the sequence presented in Figure 89 (SEQ ID NO:118) (a Vpu polypeptide); the sequence presented in Figure 63 (SEQ ID NO:92) (a Rev polypeptide); and the sequence presented in Figure 66 (SEQ ID NO:95) (a Tat polypeptide).

The native and synthetic polynucleotide sequences encoding the HIV polypeptides of the present invention typically have at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught herein. Further, in certain embodiments, the polynucleotide sequences encoding the HIV polypeptides of the invention will exhibit 100% sequence identity to the sequences taught herein.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

The present invention further includes recombinant expression systems for use in selected host cells, wherein the recombinant expression systems employ one or more of the polynucleotides and expression cassettes of the present invention. In such systems, the polynucleotide sequences are operably linked to control elements compatible with expression

in the selected host cell. Numerous expression control elements are known to those in the art, including, but not limited to, the following: transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences. Exemplary transcription promoters include, but are not limited to those derived from CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

In another aspect the invention includes cells comprising one or more of the expression cassettes of the present invention where the polynucleotide sequences are operably linked to control elements compatible with expression in the selected cell. In one embodiment such cells are mammalian cells. Exemplary mammalian cells include, but are not limited to, BHK, VERO, HT1080, 293, RD, COS-7, and CHO cells. Other cells, cell types, tissue types, etc., that may be useful in the practice of the present invention include, but are not limited to, those obtained from the following: insects (e.g., *Trichoplusia ni* (Tn5) and Sf9), bacteria, yeast, plants, antigen presenting cells (e.g., macrophage, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), primary cells, immortalized cells, tumor-derived cells.

In a further aspect, the present invention includes compositions for generating an immunological response, where the composition typically comprises at least one of the expression cassettes of the present invention and may, for example, contain combinations of expression cassettes (such as one or more expression cassettes carrying a Pol-polypeptide-encoding polynucleotide, one or more expression cassettes carrying a Gag-polypeptide-encoding polynucleotide, one or more expression cassettes carrying accessory polypeptide-encoding polynucleotides (*e.g.*, native or synthetic vpu, vpr, nef, vif, tat, rev), and/or one or more expression cassettes carrying an Env-polypeptide-encoding polynucleotide). Such compositions may further contain an adjuvant or adjuvants. The compositions may also contain one or more Type C HIV polypeptides. The Type C HIV polypetpides may correspond to the polypeptides encoded by the expression cassette(s) in the composition, or may be different from those encoded by the expression cassettes. An example of the polynucleotide in the expression cassette encoding the same polypeptide as is being provided in the composition is as follows: the polynucleotide in the expression cassette encodes the Gag-polypeptide of Figure 1 (SEQ ID NO:3), and the polypeptide (SEQ ID NO:17) is the polypeptide encoded by the sequence shown in Figure 1. An example of the polynucleotide in

the expression cassette encoding a different polypeptide as is being provided in the composition is as follows: an expression cassette having a polynucleotide encoding a Gag-polymerase polypeptide, and the polypeptide provided in the composition may be a Gag and/or Gag-protease polypeptide. In compositions containing both expression cassettes (or polynucleotides of the present invention) and polypeptides, various expression cassettes of the present invention can be mixed and/or matched with various Type C HIV polypeptides described herein.

In another aspect the present invention includes methods of immunization of a subject. In the method any of the above described compositions are into the subject under conditions that are compatible with expression of the expression cassette(s) in the subject. In one embodiment, the expression cassettes (or polynucleotides of the present invention) can be introduced using a gene delivery vector. The gene delivery vector can, for example, be a non-viral vector or a viral vector. Exemplary viral vectors include, but are not limited to Sindbis-virus derived vectors, retroviral vectors, and lentiviral vectors. Compositions useful for generating an immunological response can also be delivered using a particulate carrier. Further, such compositions can be coated on, for example, gold or tungsten particles and the coated particles delivered to the subject using, for example, a gene gun. The compositions can also be formulated as liposomes. In one embodiment of this method, the subject is a mammal and can, for example, be a human.

In a further aspect, the invention includes methods of generating an immune response in a subject. Any of the expression cassettes described herein can be expressed in a suitable cell to provide for the expression of the Type C HIV polypeptides encoded by the polynucleotides of the present invention. The polypeptide(s) are then isolated (e.g., substantially purified) and administered to the subject in an amount sufficient to elicit an immune response. In certain embodiments, the methods comprise administration of one or more of the expression cassettes or polynucleotides of the present invention, using any of the gene delivery techniques described herein. In other embodiments, the methods comprise co-administration of one or more of the expression cassettes or polynucleotides of the present invention and one or more polypeptides, wherein the polypeptides can be expressed from these polynucleotides or can be other subtype C HIV polypeptides. In other embodiments, the methods comprise co-administration of multiple expression cassettes or polynucleotides of the present invention. In still further embodiments, the methods comprise co-

administration of multiple polypeptides, for example polypeptides expressed from the polynucleotides of the present invention and/or other subtype C HIV polypeptides.

The invention further includes methods of generating an immune response in a subject, where cells of a subject are transfected with any of the above-described expression cassettes or polynucleotides of the present invention, under conditions that permit the expression of a selected polynucleotide and production of a polypeptide of interest (e.g., encoded by any expression cassette of the present invention). By this method an immunological response to the polypeptide is elicited in the subject. Transfection of the cells may be performed *ex vivo* and the transfected cells are reintroduced into the subject. Alternately, or in addition, the cells may be transfected *in vivo* in the subject. The immune response may be humoral and/or cell-mediated (cellular). In a further embodiment, this method may also include administration of an Type C HIV polypeptides before, concurrently with, and/or after introduction of the expression cassette into the subject.

These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 (SEQ ID NO:3) shows the nucleotide sequence of a polynucleotide encoding a synthetic Gag polypeptide. The nucleotide sequence shown was obtained by modifying type C strain AF110965 and include further modifications of INS.

Figure 2 (SEQ ID NO: 4) shows the nucleotide sequence of a polynucleotide encoding a synthetic Gag polypeptide. The nucleotide sequence shown was obtained by modifying type C strain AF110967 and include further modifications of INS.

Figure 3 (SEQ ID NO:9) shows the nucleotide sequence of a polynucleotide encoding a synthetic Env polypeptide. The nucleotide sequence depicts gp160 (including a signal peptide) and was obtained by modifying type C strain AF110968. The arrows indicate the positions of various regions of the polynucleotide, including the sequence encoding a signal peptide (nucleotides 1-81) (SEQ ID NO:18), a gp120 polypeptide (nucleotides 82-1512) (SEQ ID NO:6), a gp41 polypeptide (nucleotides 1513-2547) (SEQ ID NO:10), a gp140 polypeptide (nucleotides 82-2025) (SEQ ID NO:7) and a gp160 polypeptide (nucleotides 82-2547) (SEQ ID NO:8). The codons encoding the signal peptide are modified (as described herein) from the native HIV-1 signal sequence.

Figure 4 (SEQ ID NO:15) shows the nucleotide sequence of a polynucleotide encoding a synthetic Env polypeptide. The nucleotide sequence depicts gp160 (including a signal peptide) and was obtained by modifying type C strain AF110975. The arrows indicate the positions of various regions of the polynucleotide, including the sequence encoding a signal peptide (nucleotides 1-72) (SEQ ID NO:19), a gp120 polypeptide (nucleotides 73-1509) (SEQ ID NO:12), a gp41 polypeptide (nucleotides 1510-2565) (SEQ ID NO:16), a gp140 polypeptide (nucleotides 73-2022) (SEQ ID NO:13), and a gp160 polypeptide (nucleotides 73-2565) (SEQ ID NO:14). The codons encoding the signal peptide are modified (as described herein) from the native HIV-1 signal sequence.

Figure 5 shows the location of some remaining INS in synthetic Gag sequences derived from AF110965. The changes made to these sequences are boxed in the Figures. The top line depicts a codon modified sequence of Gag polypeptides from the indicated strains (SEQ ID NO:20). The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to remove further INS and correspond to the sequence depicted in Figure 1 (SEQ ID NO:3).

Figure 6 shows the location of some remaining INS in synthetic Gag sequences derived from AF110967. The changes made to these sequences are boxed in the Figures. The top line depicts a modified sequence of Gag polypeptides from the indicated strains (SEQ ID NO:21). The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to remove further INS and correspond to the sequence depicted in Figure 2 (SEQ ID NO:4).

Figure 7 is a schematic depicting the selected domains in the *Pol* region of HIV.

Figure 8 (SEQ ID NO:30) depicts the nucleotide sequence of the synthetic construct designated PR975(+). "(+)" indicates that the reverse transcriptase is functional. This construct includes sequence from p2 (nucleotides 16 to 54 of SEQ ID NO:30); p7 (nucleotides 55 to 219 of SEQ ID NO:30); p1/p6 (nucleotides 220-375 of SEQ ID NO:30); prot (nucleotides 376 to 672 of SEQ ID NO:30), reverse transcriptase (nucleotides 673 to 2352 of SEQ ID NO:30); and 6 amino acids of integrase shown in Figure 7 (nucleotides 2353 to 2370 of SEQ ID NO:30). In addition, the construct contains a multiple cloning site (MCS, nucleotides 2425 to 2463 of SEQ ID NO:30) for insertion of a transgene and a YMDD epitope cassette (nucleotides 2371 to 2424 of SEQ ID NO:30).

Figure 9 (SEQ ID NO:31) depicts the nucleotide sequence of the synthetic construct designated PR975YM. As illustrated in Figure 7,. the RT region includes a mutation in the catalytic center (mut. cat. center). "YM" refers to constructs in which the nucleotides encode the amino acids AP instead of YMDD in this region. Reverse transcriptase is not functional in this construct. This construct includes sequence from the p2 (nucleotides 16 to 54 of SEQ ID NO:31); p7 (nucleotides 55 to 219 of SEQ ID NO:31); p1/p6 (nucleotides 220 to 375 of SEQ ID NO:31); prot (nucleotides 376 to 672 of SEQ ID NO:31); and reverse transcriptase (nucleotides 673 to 2346 of SEQ ID NO:31) shown in Figure 7, although the reverse transcriptase protein is not functional. In addition, the construct contains a multiple cloning site (MCS, nucleotides 2419 to 2457 of SEQ ID NO:31) for insertion of a transgene and a YMDD epitope cassette (nucleotides 2365 to 2418 of SEQ ID NO:31).

Figure 10 (SEQ ID NO:32) depicts the nucleotide sequence of the synthetic construct designated PR975YMWM. "YM" refers to constructs in which the nucleotides encode the amino acids AP instead of YMDD in this region. "WM" refers to constructs in which the nucleotides encode amino acids PI instead of WMGY in this region. This construct includes sequence from the p2 (nucleotides 16 to 54 of SEQ ID NO:32); p7 (nucleotides 55 to 219 of SEQ ID NO:32); p1/p6 (nucleotides 220 to 375 of SEQ ID NO:32); prot (nucleotides 376 to 672 of SEQ ID NO:32); and reverse transcriptase (nucleotides 673 to 2340 of SEQ ID NO:32) shown in Figure 7, although the reverse transcriptase protein is not functional. In addition, the construct contains a multiple cloning site (MCS, nucleotides 2413 to 2451 of SEQ ID NO:32) for insertion of a transgene and a YMDD epitope cassette (nucleotides 2359 to 2412 of SEQ ID NO:32).

Figure 11 (SEQ ID NO:33) depicts the nucleotide sequence of 8_5_TV1_C.ZA. Various regions are shown in Table A.

Figure 12 (SEQ ID NO:34) depicts the wild type nucleotide sequence of AF110975 Pol from p2gag until p7gag.

Figure 13 (SEQ ID NO:35) depicts the wild type nucleotide sequence of AF110975 Pol from p1 through the first 6 amino acids of the integrase protein.

Figure 14 (SEQ ID NO:36) depicts the nucleotide sequence of a cassette encoding Ile178 through Serine 191 of reverse transcriptase.

Figure 15 (SEQ ID NO:37) shows amino acid sequence which includes an epitope in the region of the catalytic center of the reverse transcriptase protein.

Figure 16 (SEQ ID NO:45) depicts the nucleotide sequence of 12-5_1_TV2_C.ZA.

Figure 17 (SEQ ID NO:46) depicts the nucleotide sequence of a synthetic Env-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence corresponds to a short (97 base pair) common region.

Figure 18 (SEQ ID NO:47) depicts the nucleotide sequence of a synthetic Env-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence corresponds to a common region in Env.

Figure 19 (SEQ ID NO:48) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Env.

Figure 20 (SEQ ID NO:49) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 21 (SEQ ID NO:50) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Env gp160.

Figure 22 (SEQ ID NO:51) depicts the nucleotide sequence of a synthetic Gag-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 23 (SEQ ID NO:52) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Gag.

Figure 24 (SEQ ID NO:53) depicts the nucleotide sequence of a synthetic Gag-encoding polynucleotide (major homology region) derived from 8_5_TV1_C.ZA.

Figure 25 (SEQ ID NO:54) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Gag major homology region.

Figure 26 (SEQ ID NO:55) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 27 (SEQ ID NO:56) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Nef.

Figure 28 (SEQ ID NO:57) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence includes a mutation at position 125 which results in a non-functional gene product.

Figure 29 (SEQ ID NO:58) depicts the nucleotide sequence of a synthetic RNAseH-encoding polynucleotide derived from 8_5_TV1_C.ZA. RnaseH is a functional domain of the Pol gene, corresponding to p15 (Table A).

Figure 30 (SEQ ID NO:59) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA RNAseH.

Figure 31 (SEQ ID NO:60) depicts the nucleotide sequence of a synthetic integrase (Int)-encoding polynucleotide derived from 8_5_TV1_C.ZA. Int is a functional domain of the Pol gene, corresponding to p31 (Table A).

Figure 32 (SEQ ID NO:61) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Int.

Figure 33 (SEQ ID NO:62) depicts the nucleotide sequence of a synthetic Pol-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 34 (SEQ ID NO:63) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Pol.

Figure 35 (SEQ ID NO:64) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 36 (SEQ ID NO:65) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Prot.

Figure 37 (SEQ ID NO:66) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide derived from 8_5_TV1_C.ZA containing a mutation in which results in inactivation of the protease.

Figure 38 (SEQ ID NO:67) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA inactivated Prot.

Figure 39 (SEQ ID NO:68) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide and a synthetic reverse transcriptase (RT)-encoding polynucleotide, both derived from 8_5_TV1_C.ZA. The Prot and RT sequences both contain a mutation which results in inactivation of the gene product.

Figure 40 (SEQ ID NO:69) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA inactivated Prot/mutated RT.

Figure 41 (SEQ ID NO:70) depicts the nucleotide sequence of a synthetic protease (prot)-encoding polynucleotide and a synthetic reverse transcriptase (RT)-encoding polynucleotide, both derived from 8_5_TV1_C.ZA.

Figure 42 (SEQ ID NO:71) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Prot and RT.

Figure 43 (SEQ ID NO:72) depicts the nucleotide sequence of a synthetic rev-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 1 of rev. Wild-type rev has two exons.

Figure 44 (SEQ ID NO:73) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 1 of Rev.

Figure 45 (SEQ ID NO:74) depicts the nucleotide sequence of a synthetic rev-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 2 of rev.

Figure 46 (SEQ ID NO:75) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 2 of Rev.

Figure 47 (SEQ ID NO:76) depicts the nucleotide sequence of a synthetic RT-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 48 (SEQ ID NO:77) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA RT.

Figure 49 (SEQ ID NO:78) depicts the nucleotide sequence of a synthetic RT-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic polynucleotide includes a mutation in the RT coding sequence which renders the gene product inactive.

Figure 50 (SEQ ID NO:79) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA RT including a mutation which inactivates the RT gene product.

Figure 51 (SEQ ID NO:80) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 1 of Tat and further includes a mutation that renders the Tat gene product non-functional. Wild-type Tat has two exons.

Figure 52 (SEQ ID NO:81) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 1 of Tat.

Figure 53 (SEQ ID NO:82) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 1 of Tat.

Figure 54 (SEQ ID NO:83) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exon 2 of Tat.

Figure 55 (SEQ ID NO:84) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA exon 2 of Tat.

Figure 56 (SEQ ID NO:85) depicts the nucleotide sequence of a synthetic Vif-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 57 (SEQ ID NO:86) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Vif.

Figure 58 (SEQ ID NO:87) depicts the nucleotide sequence of a synthetic Vpr-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 59 (SEQ ID NO:88) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Vpr.

Figure 60 (SEQ ID NO:89) depicts the nucleotide sequence of a synthetic Vpu-encoding polynucleotide derived from 8_5_TV1_C.ZA.

Figure 61 (SEQ ID NO:90) depicts the wild-type nucleotide sequence of 8_5_TV1_C.ZA Vpu.

Figure 62 (SEQ ID NO:91) depicts the nucleotide sequence of a synthetic rev-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic sequence depicted corresponds to exons 1 and 2 of rev.

Figure 63 (SEQ ID NO:92) depicts the wild-type nucleotide sequence of exons 1 and 2 of rev derived from 8_5_TV1_C.ZA.

Figure 64 (SEQ ID NO:93) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic polynucleotide includes both exons 1 and 2 of Tat and further includes a mutation in exon 1 which renders the gene product non-functional.

Figure 65 (SEQ ID NO:94) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from 8_5_TV1_C.ZA. The synthetic polynucleotide includes both exons 1 and 2 of Tat.

Figure 66 (SEQ ID NO:95) depicts the wild-type nucleotide sequence of exons 1 and 2 of Tat derived from 8_5_TV1_C.ZA.

Figure 67 (SEQ ID NO:96) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 8_5_TV1_C.ZA. The sequence includes a mutation at position 125 which results in a non-functional gene product and a mutation that eliminates the myristoylation site of the Nef gene product.

Figure 68 (SEQ ID NO:97) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 69 (SEQ ID NO:98) depicts the wild-type nucleotide sequence of Env gp160 derived from 12-5_1_TV2_C.ZA.

Figure 70 (SEQ ID NO:99) depicts the nucleotide sequence of a synthetic Gag-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 71 (SEQ ID NO:100) depicts the wild-type nucleotide sequence of Gag derived from 12-5_1_TV2_C.ZA.

Figure 72 (SEQ ID NO:101) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 73 (SEQ ID NO:102) depicts the wild-type nucleotide sequence of Nef derived from 12-5_1_TV2_C.ZA.

Figure 74 (SEQ ID NO:103) depicts the nucleotide sequence of a synthetic Pol-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 75 (SEQ ID NO:104) depicts the wild-type nucleotide sequence of Pol derived from 12-5_1_TV2_C.ZA.

Figure 76 (SEQ ID NO:105) depicts the nucleotide sequence of a synthetic Rev-encoding polynucleotide derived from exon 1 of Rev from 12-5_1_TV2_C.ZA.

Figure 77 (SEQ ID NO:106) depicts the wild-type nucleotide sequence of exon 1 of Rev derived from 12-5_1_TV2_C.ZA.

Figure 78 (SEQ ID NO:107) depicts the nucleotide sequence of a synthetic Rev-encoding polynucleotide derived from exon 2 of Rev from 12-5_1_TV2_C.ZA.

Figure 79 (SEQ ID NO:108) depicts the wild-type nucleotide sequence of exon 2 of Rev derived from 12-5_1_TV2_C.ZA.

Figure 80 (SEQ ID NO:109) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from exon 1 of Tat from 12-5_1_TV2_C.ZA.

Figure 81 (SEQ ID NO:110) depicts the wild-type nucleotide sequence of exon 1 of Tat derived from 12-5_1_TV2_C.ZA.

Figure 82 (SEQ ID NO:111) depicts the nucleotide sequence of a synthetic Tat-encoding polynucleotide derived from exon 2 of Tat from 12-5_1_TV2_C.ZA.

Figure 83 (SEQ ID NO:112) depicts the wild-type nucleotide sequence of exon 2 of Tat derived from 12-5_1_TV2_C.ZA.

Figure 84 (SEQ ID NO:113) depicts the nucleotide sequence of a synthetic Vif-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 85 (SEQ ID NO:114) depicts the wild-type nucleotide sequence of Vif derived from 12-5_1_TV2_C.ZA.

Figure 86 (SEQ ID NO:115) depicts the nucleotide sequence of a synthetic Vpr-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 87 (SEQ ID NO:116) depicts the wild-type nucleotide sequence of Vpr derived from 12-5_1_TV2_C.ZA.

Figure 88 (SEQ ID NO:117) depicts the nucleotide sequence of a synthetic Vpu-encoding polynucleotide derived from 12-5_1_TV2_C.ZA.

Figure 89 (SEQ ID NO:118) depicts the wild-type nucleotide sequence of Vpu derived from 12-5_1_TV2_C.ZA.

Figure 90 (SEQ ID NO:119) depicts the nucleotide sequence of a synthetic Env gp120-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a codon modified signal peptide leader sequence (nucleotides 7 to 87); a gp120 coding sequence (nucleotides 88 to 1464); a stop codon (nucleotides 1465 to 1467); an XhoI restriction site (nucleotides 1468 to 1473).

Figure 91 (SEQ ID NO:120) depicts the nucleotide sequence of a synthetic Env gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp140 coding sequence (nucleotides 88 to 1977); a stop codon (nucleotides 1978 to 1980); an XhoI restriction site (nucleotides 1981 to 1986).

Figure 92 (SEQ ID NO:121) depicts the nucleotide sequence of a synthetic Env gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted and the sequence includes mutations in the cleavage site that prevent the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); gp140 coding sequence (nucleotides 88 to 1977); a stop codon (nucleotides 1978 to 1980); an XhoI restriction site (nucleotides 1981 to 1986).

Figure 93 (SEQ ID NO:122) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V1/V2 regions are deleted. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); gp160 coding sequence (nucleotides 88 to 2388); a stop codon (nucleotides 2389 to 2391); an XhoI restriction site (nucleotides 2392 to 2397).

Figure 94 (SEQ ID NO:123) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 88 to 2520); a stop codon (nucleotides 2521 to 2523); an XhoI restriction site (nucleotides 2524 to 2529).

Figure 95 (SEQ ID NO:124) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The V2 region is deleted and the cleavage site is mutated. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 88 to 2520); a stop codon (nucleotides 2521 to 2523); an XhoI restriction site (nucleotides 2524 to 2529).

Figure 96 (SEQ ID NO:125) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes a TPA1 leader sequence. The sequence includes: a SalI restriction site (nucleotides 1 to 6); a Kozak sequence (nucleotides 7 to 12); a TPA1 signal peptide leader sequence (nucleotides 13 to 87); a gp160 coding sequence (nucleotides 88 to 2604); a stop codon (nucleotides 2605 to 2607); an XhoI restriction site (nucleotides 2608 to 2613).

Figure 97 (SEQ ID NO:126) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The sequence includes: an EcoRI restriction site (nucleotides 1 to 6); a modified signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 8 to 2607); a stop codon (nucleotides 2608 to 2610); an XhoI restriction site (nucleotides 2611 to 2616).

Figure 98 (SEQ ID NO:127) depicts the nucleotide sequence of a synthetic Env gp160-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes a wild type leader sequence. The sequence includes: an EcoRI restriction site

(nucleotides 1 to 6); a native (unmodified) signal peptide leader sequence (nucleotides 7 to 87); a gp160 coding sequence (nucleotides 88 to 2607); a stop codon (nucleotides 2608 to 2610); an XhoI restriction site (nucleotides 2611 to 2616).

Figure 99 (SEQ ID NO:128) depicts the nucleotide sequence of wild type gp160 derived from 8_2_TV1_C.ZA.

Figure 100 (SEQ ID NO:131) depicts the nucleotide sequence of a synthetic Env gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes a TPA1 leader sequence (nucleotides 1-75); a gp140 coding sequence (nucleotides 76 to 2049); a stop codon (nucleotides 2050 to 2052)

Figure 101 (SEQ ID NO:132) depicts the nucleotide sequence of a synthetic gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes an EcoRI restriction site (nucleotides 1 to 6); a leader sequence modified from the TV1_C.ZA wild-type leader sequence (nucleotides 7 to 87); a gp140 coding sequence (nucleotides 88 to 2064); a stop codon (nucleotides 2065 to 2067); a XhoI restriction site (nucleotides 2068 to 2073).

Figure 102 (SEQ ID NO:133) depicts the nucleotide sequence of a synthetic gp140-encoding polynucleotide derived from 8_2_TV1_C.ZA. The nucleotide sequence includes wild-type TV1_C.ZA unmodified leader sequence. The nucleotide sequence includes a restriction site (nucleotides 1 to 6); a wild type leader sequence (nucleotides 7 to 87); a gp140 coding sequence (nucleotides 88 to 2064); a stop codon (nucleotides 2065 to 2067); a XhoI restriction site (nucleotides 2068-2073).

Figure 103 (SEQ ID NO:134) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 12-5_1_TV2_C.ZA. The sequence includes a mutation at position 125 which results in a non-functional gene product.

Figure 104 (SEQ ID NO:135) depicts the nucleotide sequence of a synthetic Nef-encoding polynucleotide derived from 12-5_1_TV2_C.ZA. The synthetic polynucleotide includes a mutation that eliminates the myristoylation site of the Nef gene product.

Figure 105 depicts an alignment of Env polypeptides from various HIV isolates. The regions between the arrows indicate regions (of TV1 and TV2 clones) in the beta and/or bridging sheet region(s) that can be deleted and/or truncated. The "*" denotes N-linked glycosylation sites (of TV1 and TV2 clones), one or more of which can be modified (*e.g.*, deleted and/or mutated).

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

1. DEFINITIONS

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

"Synthetic" sequences, as used herein, refers to Type C HIV polypeptide-*encoding* polynucleotides whose expression has been modified as described herein, for example, by codon substitution and inactivation of inhibitory sequences. "Wild-type" or "native" sequences, as used herein, refers to polypeptide encoding sequences that are essentially as they are found in nature, e.g., Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env and/or Nef encoding sequences as found in Type C isolates, *e.g.*, AF110965, AF110967, AF110968, AF110975, 8_5_TV1_C.ZA, 8_2_TV1_C.ZA or 12-5_1_TV2_C.ZA. The various regions of the HIV genome are shown in Table A, with numbering relative to 8_5_TV1_C.ZA (SEQ ID NO:33). Thus, the term "Pol" refers to one or more of the following polypeptides: polymerase (p6Pol); protease (prot); reverse transcriptase (p66RT or RT); RNAseH (p15RNAseH); and/or integrase (p31Int or Int).

As used herein, the term "virus-like particle" or "VLP" refers to a nonreplicating, viral ıell, derived from any of several viruses discussed further below. VLPs are ɜnerally composed of one or more viral proteins, such as, but not limited to those proteins ɔferred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming olypeptides derived from these proteins. VLPs can form spontaneously upon recombinant xpression of the protein in an appropriate expression system. Methods for producing articular VLPs are known in the art and discussed more fully below. The presence of VLPs ɔllowing recombinant expression of viral proteins can be detected using conventional ɛchniques known in the art, such as by electron microscopy, X-ray crystallography, and the ike. See, e.g., Baker et al., *Biophys. J.* (1991) 60:1445-1456; Hagensee et al., *J. Virol.* 1994) 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation nd/or identified by characteristic density banding. Alternatively, cryoelectron microscopy an be performed on vitrified aqueous samples of the VLP preparation in question, and mages recorded under appropriate exposure conditions.

By "particle-forming polypeptide" derived from a particular viral protein is meant a ull-length or near full-length viral protein, as well as a fragment thereof, or a viral protein vith internal deletions, which has the ability to form VLPs under conditions that favor VLP ormation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, runcated and partial sequences, as well as analogs and precursor forms of the reference nolecule. The term therefore intends deletions, additions and substitutions to the sequence, ɜo long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also includes deletions, additions and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP. Preferred substitutions are those which are conservative in nature, i.e., those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cystine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids.

An "antigen" refers to a molecule containing one or more epitopes (either linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen." Normally, a B-cell epitope will include at least about 5 amino acids but can be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a helper T-cell epitope at least about 12-20 amino acids. Normally, an epitope will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (i.e., antigens which are separate and discrete from a whole organism with which the antigen is associated in nature), as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes. Antibodies such as anti-idiotype antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or polynucleotide which expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the definition of antigen herein.

For purposes of the present invention, antigens can be derived from any of several known viruses, bacteria, parasites and fungi, as described more fully below. The term also intends any of the various tumor antigens. Furthermore, for purposes of the present invention, an "antigen" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the ability to elicit an immunological response, as defined herein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the antigens.

An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTL"s). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the

surfaces of cells. CTLs help induce and promote the destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376. Recent methods of measuring cell-mediated immune response include measurement of intracellular cytokines or cytokine secretion by T-cell populations, or by measurement of epitope specific T-cells (e.g., by the tetramer technique)(reviewed by McMichael, A.J., and O'Callaghan, C.A., *J. Exp. Med.* **187(9)**1367-1371, 1998; Mcheyzer-Williams, M.G., et al, *Immunol. Rev.* **150**:5-21, 1996; Lalvani, A., et al, *J. Exp. Med.* **186**:859-865, 1997).

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or γδ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent

cell cytotoxicity (ADCC) to provide protection to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

An "immunogenic composition" is a composition that comprises an antigenic molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest. The immunogenic composition can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal (*e.g.*, intra-rectally or intra-vaginally) administration.

By "subunit vaccine" is meant a vaccine composition which includes one or more selected antigens but not all antigens, derived from or homologous to, an antigen from a pathogen of interest such as from a virus, bacterium, parasite or fungus. Such a composition is substantially free of intact pathogen cells or pathogenic particles, or the lysate of such cells or particles. Thus, a "subunit vaccine" can be prepared from at least partially purified (preferably substantially purified) immunogenic polypeptides from the pathogen, or analogs thereof. The method of obtaining an antigen included in the subunit vaccine can thus include standard purification techniques, recombinant production, or synthetic production.

"Substantially purified" general refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a substantially purified component comprises 50%, preferably 80%-85%, more preferably 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence such as a stop codon may be located 3' to the coding sequence.

Typical "control elements", include, but are not limited to, transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences (located 3' to the translation stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), and translation termination sequences.

A "polynucleotide coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. Exemplary coding sequences are the modified viral polypeptide-coding sequences of the present invention. A transcription termination sequence may be located 3' to the coding sequence. Typical "control elements", include, but are not limited to, transcription regulators, such as promoters, transcription enhancer elements, transcription termination signals, and polyadenylation sequences; and translation regulators, such as sequences for optimization of initiation of translation, *e.g.*, Shine-Dalgarno (ribosome binding site) sequences, Kozak sequences (i.e., sequences for the optimization of translation, located, for example, 5' to the coding sequence), leader sequences, translation initiation codon (e.g., ATG), and translation termination sequences. In certain embodiments, one or more translation regulation or initiation sequences (*e.g.*, the leader sequence) are derived from wild-type translation initiation sequences, *i.e.*, sequences that regulate translation of the coding region in their native state. Wild-type leader sequences that have been modified, using the methods described herein, also find use in the present invention. Promoters can include inducible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), repressible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), and constitutive promoters.

A "nucleic acid" molecule can include, but is not limited to, procaryotic sequences, eucaryotic mRNA, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter

operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting procaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then can be determined between the compared polypeptide sequences. Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of the mRNA for that gene (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to

an exact nucleotide to nucleotide or amino acid to amino acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be extended to use with peptide sequences using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program BLAST, which can also be used with default parameters. For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand =

both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address: http://www.ncbi.nlm.gov/cgi-bin/BLAST.

One of skill in the art can readily determine the proper search parameters to use for a given sequence, exemplary preferred Smith Waterman based parameters are presented above. For example, the search parameters may vary based on the size of the sequence in question. Thus, for the polynucleotide sequences of the present invention the length of the polynucleotide sequence disclosed herein is searched against a selected database and compared to sequences of essentially the same length to determine percent identity. For example, a representative embodiment of the present invention would include an isolated polynucleotide having X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least about a selected level of percent identity relative to Y contiguous nucleotides of the sequences described herein, and (ii) for search purposes X equals Y, wherein Y is a selected reference polynucleotide of defined length.

The sequences of the present invention can include fragments of the sequences, for example, from about 15 nucleotides up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing, Figures, and claims), including all integer values falling within the above-described range. For example, fragments of the polynucleotide sequences of the present invention may be 30-60 nucleotides, 60-120 nucleotides, 120-240 nucleotides, 240-480 nucleotides, 480-1000 nucleotides, and all integer values therebetween.

The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%, greater than 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% up to 100% (including all integer values falling within these described ranges) sequence identity to the synthetic expression cassette (and purified polynucleotide) sequences disclosed herein (for example, to the claimed sequences or other sequences of the present invention) when the sequences of the present invention are used as the query sequence against, for example, a database of sequences.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the

efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *supra* or Ausubel et al., *supra*). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions.

f a particular set of hybridization conditions is selected following standard
art (see, for example, Sambrook, et al., *supra* or Ausubel et al., *supra*).
olynucleotide is "derived from" second polynucleotide if it has the same or
e same basepair sequence as a region of the second polynucleotide, its cDNA,
ereof, or if it displays sequence identity as described above.
olypeptide is "derived from" a second polypeptide if it is (i) encoded by a first
derived from a second polynucleotide, or (ii) displays sequence identity to the
tides as described above.
lly, a viral polypeptide is "derived from" a particular polypeptide of a virus
ide) if it is (i) encoded by an open reading frame of a polynucleotide of that
ynucleotide), or (ii) displays sequence identity to polypeptides of that virus as
e.
led by" refers to a nucleic acid sequence which codes for a polypeptide
rein the polypeptide sequence or a portion thereof contains an amino acid
least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even
y at least 15 to 20 amino acids from a polypeptide encoded by the nucleic acid
o encompassed are polypeptide sequences which are immunologically
th a polypeptide encoded by the sequence. Further, polyproteins can be
fusing in-frame two or more polynucleotide sequences encoding polypeptide
ducts. Further, polycistronic coding sequences may be produced by placing
olynucleotide sequences encoding polypeptide products adjacent each other,
r the control of one promoter, wherein each polypeptide coding sequence may
include sequences for internal ribosome binding sites.
ied polynucleotide" refers to a polynucleotide of interest or fragment thereof
tially free, e.g., contains less than about 50%, preferably less than about 70%,
ferably less than about 90%, of the protein with which the polynucleotide is
ciated. Techniques for purifying polynucleotides of interest are well-known in
clude, for example, disruption of the cell containing the polynucleotide with a
ent and separation of the polynucleotide(s) and proteins by ion-exchange
hy, affinity chromatography and sedimentation according to density.
ucleic acid immunization" is meant the introduction of a nucleic acid molecule
or more selected antigens into a host cell, for the *in vivo* expression of an

antigen, antigens, an epitope, or epitopes. The nucleic acid molecule can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter case, the transformed cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

"Gene transfer" or "gene delivery" refers to methods or systems for reliably inserting DNA of interest into a host cell. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred replicons (e.g., episomes), or integration of transferred genetic material into the genomic DNA of host cells. Gene delivery expression vectors include, but are not limited to, vectors derived from alphaviruses, pox viruses and vaccinia viruses. When used for immunization, such gene delivery expression vectors may be referred to as vaccines or vaccine vectors.

"T lymphocytes" or "T cells" are non-antibody producing lymphocytes that constitute a part of the cell-mediated arm of the immune system. T cells arise from immature lymphocytes that migrate from the bone marrow to the thymus, where they undergo a maturation process under the direction of thymic hormones. Here, the mature lymphocytes rapidly divide increasing to very large numbers. The maturing T cells become immunocompetent based on their ability to recognize and bind a specific antigen. Activation of immunocompetent T cells is triggered when an antigen binds to the lymphocyte's surface receptors.

The term "transfection" is used to refer to the uptake of foreign DNA by a cell. A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. *See, e.g.*, Graham et al. (1973) *Virology*, 52:456, Sambrook et al. (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, and Chu et al. (1981) *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells. The term refers to both stable and transient uptake of the genetic material, and includes uptake of peptide- or antibody-linked DNAs.

A "vector" is capable of transferring gene sequences to target cells (e.g., viral vectors, non-viral vectors, particulate carriers, and liposomes). Typically, "vector construct,"

"expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Transfer of a "suicide gene" (e.g., a drug-susceptibility gene) to a target cell renders the cell sensitive to compounds or compositions that are relatively nontoxic to normal cells. Moolten, F.L. (1994) *Cancer Gene Ther.* 1:279-287. Examples of suicide genes are thymidine kinase of herpes simplex virus (HSV-tk), cytochrome P450 (Manome et al. (1996) *Gene Therapy* 3:513-520), human deoxycytidine kinase (Manome et al. (1996) *Nature Medicine* 2(5):567-573) and the bacterial enzyme cytosine deaminase (Dong et al. (1996) *Human Gene Therapy* 7:713-720). Cells which express these genes are rendered sensitive to the effects of the relatively nontoxic prodrugs ganciclovir (HSV-tk), cyclophosphamide (cytochrome P450 2B1), cytosine arabinoside (human deoxycytidine kinase) or 5-fluorocytosine (bacterial cytosine deaminase). Culver et al. (1992) *Science* 256:1550-1552, Huber et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:8302-8306.

A "selectable marker" or "reporter marker" refers to a nucleotide sequence included in a gene transfer vector that has no therapeutic activity, but rather is included to allow for simpler preparation, manufacturing, characterization or testing of the gene transfer vector.

A "specific binding agent" refers to a member of a specific binding pair of molecules wherein one of the molecules specifically binds to the second molecule through chemical and/or physical means. One example of a specific binding agent is an antibody directed against a selected antigen.

By "subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The system described above is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

By "pharmaceutically acceptable" or "pharmacologically acceptable" is meant a material which is not biologically or otherwise undesirable, i.e., the material may be

administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the composition in which it is contained.

By "physiological pH" or a "pH in the physiological range" is meant a pH in the range of approximately 7.2 to 8.0 inclusive, more typically in the range of approximately 7.2 to 7.6 inclusive.

As used herein, "treatment" refers to any of (I) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may be effected prophylactically (prior to infection) or therapeutically (following infection).

By "co-administration" is meant administration of more than one composition or molecule. Thus, co-administration includes concurrent administration or sequentially administration (in any order), via the same or different routes of administration. Non-limiting examples of co-administration regimes include, co-administration of nucleic acid and polypeptide; co-administration of different nucleic acids (*e.g.*, different expression cassettes as described herein and/or different gene delivery vectors); and co-administration of different polypeptides (*e.g.*, different HIV polypeptides and/or different adjuvants). The term also encompasses multiple administrations of one of the co-administered molecules or compositions (*e.g.*, multiple administrations of one or more of the expression cassettes described herein followed by one or more administrations of a polypeptide-containing composition). In cases where the molecules or compositions are delivered sequentially, the time between each administration can be readily determined by one of skill in the art in view of the teachings herein.

"Lentiviral vector", and "recombinant lentiviral vector" refer to a nucleic acid construct which carries, and within certain embodiments, is capable of directing the expression of a nucleic acid molecule of interest. The lentiviral vector include at least one transcriptional promoter/enhancer or locus defining element(s), or other elements which control gene expression by other means such as alternate splicing, nuclear RNA export, post-translational modification of messenger, or post-transcriptional modification of protein. Such vector constructs must also include a packaging signal, long terminal repeats (LTRS) or portion thereof, and positive and negative strand primer binding sites appropriate to the retrovirus used (if these are not already present in the retroviral vector). Optionally, the

recombinant lentiviral vector may also include a signal which directs polyadenylation, selectable markers such as Neo, TK, hygromycin, phleomycin, histidinol, or DHFR, as well as one or more restriction sites and a translation termination sequence. By way of example, such vectors typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second strand DNA synthesis, and a 3'LTR or a portion thereof

"Lentiviral vector particle" as utilized within the present invention refers to a lentivirus which carries at least one gene of interest. The retrovirus may also contain a selectable marker. The recombinant lentivirus is capable of reverse transcribing its genetic material (RNA) into DNA and incorporating this genetic material into a host cell's DNA upon infection. Lentiviral vector particles may have a lentiviral envelope, a non-lentiviral envelope (e.g., an ampho or VSV-G envelope), or a chimeric envelope.

"Nucleic acid expression vector" or "Expression cassette" refers to an assembly which is capable of directing the expression of a sequence or gene of interest. The nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well. Expression cassettes described herein may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include a bacterial origin of replication, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA *(e.g.,* a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

"Packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant retrovirus which are lacking in a recombinant retroviral vector. Typically, such packaging cells contain one or more expression cassettes which are capable of expressing proteins which encode *Gag, pol and* env proteins.

"Producer cell" or "vector producing cell" refers to a cell which contains all elements necessary for production of recombinant retroviral vector particles.

2. Modes of Carrying Out the Invention

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

2.1. THE HIV GENOME

The HIV genome and various polypeptide-encoding regions are shown in Table A. The nucleotide positions are given relative to 8_5_TV1_C.ZA (SEQ ID NO:33, Figure 11). However, it will be readily apparent to one of ordinary skill in the art in view of the teachings of the present disclosure how to determine corresponding regions in other HIV strains or variants (*e.g.*, isolates HIV_{IIIb}, HIV_{SF2}, $HIV\text{-}1_{SF162}$, $HIV\text{-}1_{SF170}$, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, $HIV\text{-}1_{CM235}$,, $HIV\text{-}1_{US4}$, other HIV-1 strains from diverse subtypes(e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., $HIV\text{-}2_{UC1}$ and $HIV\text{-}2_{UC2}$), and simian immunodeficiency virus (SIV). (See, e.g., Virology, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA; for a description of these and other related viruses), using for example, sequence comparison programs (*e.g.*, BLAST and others described herein) or identification and alignment of structural features (*e.g.*, a program such as the "ALB" program described herein that can identify the various regions).

Table A: Regions of the HIV Genome relative to 8_5_TV1_C.ZA

Region	Position in nucleotide sequence
5'LTR	**1-636**
U3	1-457
R	458-553
U5	554-636
NFkB II	340-348
NFkB I	354-362
Sp1 III	379-388
Sp1 II	390-398
Sp1 I	400-410
TATA Box	429-433
TAR	474-499
Poly A signal	529-534
PBS	**638-655**
p7 binding region, packaging signal	**685-791**
Gag:	**792-2285**
p17	792-1178
p24	1179-1871
Cyclophilin A bdg.	1395-1505
MHR	1632-1694
p2	1872-1907
p7	1908-2072
Frameshift slip	2072-2078
p1	2073-2120
p6Gag	2121-2285
Zn-motif I	1950-1991
Zn-motif II	2013-2054

Pol:	**2072-5086**
p6Pol	2072-2245
Prot	2246-2542
p66RT	2543-4210
p15RNaseH	3857-4210
p31Int	4211-5086
Vif:	**5034-5612**
Hydrophilic region	5292-5315
Vpr:	**5552-5839**
Oligomerization	5552-5677
Amphipathic α-helix	5597-5653
Tat:	**5823-6038 and 8417-8509**
Tat-1 exon	5823-6038
Tat-2 exon	8417-8509
N-terminal domain	5823-5885
Trans-activation domain	5886-5933
Transduction domain	5961-5993
Rev:	**5962-6037 and 8416-8663**
Rev-1 exon	5962-6037
Rev-2 exon	8416-8663
High-affinity bdg. site	8439-8486
Leu-rich effector domain	8562-8588
Vpu:	**6060-6326**
Transmembrane domain	6060-6161
Cytoplasmic domain	6162-6326

Env (gp160):	**6244-8853**
Signal peptide	6244-6324
gp120	6325-7794
V1	6628-6729
V2	6727-6852
V3	7150-7254
V4	7411-7506
V5	7663-7674
C1	6325-6627
C2	6853-7149
C3	7255-7410
C4	7507-7662
C5	7675-7794
CD4 binding	7540-7566
gp41	7795-8853
Fusion peptide	7789-7842
Oligomerization domain	7924-7959
N-terminal heptad repeat	7921-8028
C-terminal heptad repeat	8173-8280
Immunodominant region	8023-8076
Nef:	**8855-9478**
Myristoylation	8858-8875
SH3 binding	9062-9091
Polypurine tract	9128-9154
SH3 binding	9296-9307

It will be readily apparent that one of skill in the art can readily align any sequence to that shown in Table A to determine relative locations of any particular HIV gene. For example, using one of the alignment programs described herein (*e.g.*, BLAST), other HIV Type C sequences can be aligned with 8_5_TV1_C.ZA (Table A) and locations of genes determined.

Polypeptide sequences can be similarly aligned. For example, Figure 103 shows the alignment of Env polypeptide sequences from various strains, relative to SF-162. As described in detail in co-owned WO/39303, Env polypeptides (*e.g.*, gp120, gp140 and gp160) include a "bridging sheet" comprised of 4 anti-parallel β-strands (β-2, β-3, β-20 and β-21) that form a β-sheet. Extruding from one pair of the β-strands (β-2 and β-3) are two loops, V1

and V2. The β-2 sheet occurs at approximately amino acid residue 113 (Cys) to amino acid residue 117 (Thr) while β-3 occurs at approximately amino acid residue 192 (Ser) to amino acid residue 194 (Ile), relative to SF-162 (see, Figure 103). The "V1/V2 region" occurs at approximately amino acid positions 120 (Cys) to residue 189 (Cys), relative to SF-162. Extruding from the second pair of β-strands (β-20 and β-21) is a "small-loop" structure, also referred to herein as "the bridging sheet small loop." The locations of both the small loop and bridging sheet small loop can be determined relative to HXB-2 following the teachings herein and in WO/39303. Also shown by arrows in Figure 103A-C are approximate sites for deletions sequence from the beta sheet region. The "*" denotes N-glycosylation sites that can be mutated following the teachings of the present specification.

2.2 Synthetic Expression Cassettes

2.2.1 Modification of HIV-1-Type C *Pol-, Prot-, Rt-, Int-, Gag, Env, Tat, Rev, Nef, RnaseH, Vif, Vpr,* and *Vpu* Nucleic Acid Coding Sequences

One aspect of the present invention is the generation of HIV-1 type C coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequences.

2.2.1.1. Modification of *Gag* Nucleic Acid Coding Sequences

An exemplary embodiment of the present invention is illustrated herein by modifying the Gag protein wild-type sequences obtained from the AF110965 and AF110967 strains of HIV-1, subtype C. (see, for example, Korber et al. (1998)*Human Retroviruses and Aids,* Los Alamos, New Mexico: Los Alamos National Laboratory;
Novitsky et al. (1999) *J. Virol.* 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). Also illustrated herein is the modification of wild-type sequences from novel isolates 8_5_TV1_C.ZA (also called TV001 or TV1) and 12-5_1_TV2_C.ZA (also called TV002 or TV2). SEQ ID NO:52 shows the wild-type sequence of Gag from 8_5_TV1_C.ZA and SEQ ID NO:54 shows the wild-type sequence of the major homology region of Gag (nucleotides 1632-1694 of Table A) of the same strain. SEQ ID NO:100 shows the wild-type sequence of Gag of 12-5_1_TV2_C.ZA.

Gag sequence obtained from other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Gag protein encoding sequences obtained from the isolates of HIV-1 Type C, for example as described in Novitsky et al., (1999), *supra*; Myers et al., *infra;* Virology, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA and on the World Wide Web (Internet), for example at http://hiv-web.lanl.gov/cgi-bin/hivDB3/public/wdb/ssampublic and http://hiv-web.lanl.gov.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Gag coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag coding sequences. The RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements can be inactivated by introducing multiple point mutations that do not alter the reading frame of the encoded proteins.

Subtype C Gag-encoding sequences having inactivated RRE sites are shown, for example, in Figures 1 (SEQ ID NO:3), 2 (SEQ ID NO:4), 5 (SEQ ID NO:20) and 6 (SEQ ID NO:26). Similarly, other synthetic polynucleotides derived from other Subtype C strains can be modified to inactivate the RRE sites.

Modification of the Gag polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Further, expression of the sequences results in production of virus-like particles (VLPs) by these cell lines (see below).

2.2.1.2 MODIFICATION OF *ENV* NUCLEIC ACID CODING SEQUENCES

Similarly, the present invention also includes synthetic Env-encoding polynucleotides and modified Env proteins. Wild-type Env sequences are obtained from the AF110968 and AF110975 strains as well as novel strains 8_5_TV1_C.ZA (SEQ ID NO:33) and 12-5_1_TV2_C.ZA (SEQ ID NO:45) of HIV-1, type C. (see, for example, Novitsky et al. (1999) *J. Virol.* 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). Wild-type Env sequences of 8_5_TV1_C.ZA are shown, for example, in SEQ ID NO:48 (wild-type Env common region, nucleotides 7486-7629 as shown in Table A); and SEQ ID NO:50 (wild type gp160, nucleotides 6244-8853 as shown in Table A). Wild-type Env gp160 of 12-5_1_TV2_C.ZA is shown in SEQ ID NO:98. It will be readily apparent from the disclosure herein that polynucleotides encoding fragments of Env gp160 (*e.g.*, gp120, gp41, gp140) can be readily obtained from the larger, full-length sequences disclosed herein. It will also be readily apparent that other modifications can be made, for example deletion of regions such as the V1 and/or V2 region; mutation of the cleavage site and the like (see, Example 1). Exemplary sequences of such modification as shown in SEQ ID NO:119 through 127.

Further, Env sequences obtained from other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Env protein encoding sequences obtained from the isolates of HIV-1 Type C, described above.

The codon usage pattern for Env was modified as described above for Gag so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. Experiments performed in support of the present invention show that the synthetic Env sequences were capable of higher level of protein production relative to the native Env sequences.

Modification of the Env polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Similar Env polypeptide coding sequences can be obtained, modified and tested for improved expression from a variety of isolates, including those described above for Gag.

Further modifications of Env include, but are not limited to, generating polynucleotides that encode Env polypeptides having mutations and/or deletions therein. For

instance, the hypervariable regions, V1 and/or V2, can be deleted as described herein. Additionally, other modifications, for example to the bridging sheet region and/or to N-glycosylation sites within Env can also be performed following the teachings of the present specification. (see, Figure 103A-C and WO/39303). Various combinations of these modifications can be employed to generate synthetic expression cassettes as described herein.

2.2.1.3 Modification of Sequences Including HIV-1 *Pol* Nucleic Acid Coding Sequences

The present invention also includes expression cassettes which include synthetic Pol sequences. As noted above, "Pol" includes, but is not limited to, the protein-encoding regions shown in Figure 7, for example polymerase, protease, reverse transcriptase and/or integrase-containing sequences. The regions shown in Figure 7 are described, for example, in Wan et al (1996) *Biochem. J.* 316:569-573; Kohl et al. (1988) *PNAS USA* 85:4686-4690; Krausslich et al. (1988) *J. Virol.* 62:4393-4397; Coffin, "Retroviridae and their Replication" in Virology, pp1437-1500 (Raven, New York, 1990); Patel et. al. (1995) *Biochemistry* 34:5351-5363. Thus, the synthetic expression cassettes exemplified herein include one or more of these regions and one or more changes to the resulting amino acid sequences.

Wild type Pol sequences were obtained from the AF110975, 8_5_TV1_C.ZA and 12-5_1_TV2_C.ZA strains of HIV-1, type C. (see, for example, Novitsky et al. (1999) *J. Virol.* 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). SEQ ID NO:34 shows the wild type sequence of AF110975 from the p2 through p7 region of Pol (see, Figure 7 and Table A). SEQ ID NO:35 shows the wild type sequence of AF110975 from p1 through the first 6 amino acids of integrase (see, Figure 7 and Table A). SEQ ID NO:63 and SEQ ID NO:104 show wild-type sequences of Pol from 8_5_TV1_C.ZA and 12-5_1_TV2_C.ZA, respectively (see, also, Table A).

Sequence obtained from other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Pol protein encoding sequences obtained from the isolates of HIV-1 Type C described herein.

The codon usage pattern for Pol was modified as described above for Gag and Env so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes.

Table B shows the nucleotide positions of various regions found in the Pol constructs exemplified herein (*e.g.*, SEQ ID NOs: 30-32).

Table B

Region	Position in nucleotide sequence in construct		
	PR975(+) Seq Id No:30	PR975YM Seq Id No:31	PR975(+) YMWM Seq Id No:32
Sal 1 restriction site	1-6	1-6	1-6
Kozak start codon	7-16	7-16	7-16
p2	16-54	16-54	16-54
p7	55-219	55-219	55-219
p1/p6 pol	220-375	220-375	220-375
Insertion mutation for in frame	225	225	225
p10Protease	376-672	376-672	376-672
p66RT	673-2352	673-2346	673-2340
p51RT	673-1992	673-1986	673-1980
p15RNaseH	1993-2352	1993-2346	1993-2340
catalytic center region (YMDD)	1219-1230	1219-1224	1219-1224
primer grip region (WMGY)	1357-1368	1351-1362	1351-1356
6aa Integrase	2353-2370	2347-2364	2341-2358
YMDD epitope cassette (incl. 5'+3'Gly)	2371-2424	2365-2418	2359-2412
MCS (multiple cloning site)	2425-2463	2419-2457	2413-2451
EcoR 1 restriction site	2464-2469	2458-2463	2452-2457

As shown in Table B, exemplary constructs were modified in various ways. For example, the expression constructs exemplified herein include sequence that encodes the first 6 amino acids of the integrase polypeptide. This 6 amino acid region is believed to provide a cleavage recognition site recognized by HIV protease (*see, e.g.*, McCornack et al. (1997) *FEBS Letts* 414:84-88). As noted above, certain constructs exemplified herein include a multiple cloning site (MCS) for insertion of one or more transgenes, typically at the 3' end of the construct. In addition, a cassette encoding a catalytic center epitope derived from the catalytic center in RT is typically included 3' of the sequence encoding 6 amino acids of integrase. This cassette (SEQ ID NO:36) encodes Ile178 through Serine 191 of RT (amino acids 3 through 16 of SEQ ID NO:37) and was added to keep this well conserved region as a possible CTL epitope. Further, the constructs contain an insertion mutations (position 225 of SEQ ID NOs:30 to 32) to preserve the reading frame. (see, e.g., Park et al. (1991) *J. Virol.* 65:5111).

In certain embodiments, the catalytic center and/or primer grip region of RT are modified. The catalytic center and primer grip regions of RT are described, for example, in Patel et al. (1995) *Biochem.* 34:5351 and Palaniappan et al. (1997) *J. Biol. Chem.* 272(17):11157. For example, in the construct designated PR975YM (SEQ ID NO:31), wild type sequence encoding the amino acids YMDD at positions 183-185 of p66 RT, numbered relative to AF110975, are replaced with sequence encoding the amino acids "AP". In the construct designated PR975YMWM (SEQ ID NO:32), the same mutation in YMDD is made and, in addition, the primer grip region (amino acids WMGY, residues 229-232 of p66RT, numbered relative to AF110975) are replaced with sequence encoding the amino acids "PI."

For the Pol sequence, the changes in codon usage are typically restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame; however, regions within the frameshift translation region can be modified as well. Finally, inhibitory (or instability) elements (INS) located within the coding sequences of the protease polypeptide coding sequence can be altered as well.

Experiments can be performed in support of the present invention to show that the synthetic Pol sequences were capable of higher level of protein production relative to the native Pol sequences. Modification of the Pol polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Similar Pol polypeptide coding sequences can be obtained, modified and tested for improved expression from a variety of isolates, including those described above for Gag and Env.

2.2.1.4 Modification of Other HIV Sequences

The present invention also includes expression cassettes which include synthetic HIV Type C sequences derived HIV genes other than Gag, Env and Pol, including but not limited to, regions within Gag, Env, Pol, as well as, vif, vpr, tat, rev, vpu, and nef, for example from 8_5_TV1_C.ZA (SEQ ID NO:33) or 12-5_1_TV2_C.ZA (SEQ ID NO:45). Sequences obtained from other strains can be manipulated in similar fashion following the teachings of the present specification.

As noted above, the codon usage pattern is modified as described above for Gag, Env and Pol so that the resulting nucleic acid coding sequence is comparable to codon usage found in highly expressed human genes. Experiments can be performed in support of the present invention to show that these synthetic sequences were capable of higher level of protein production relative to the native sequences and that modification of the wild-type polypeptide coding sequences results in improved expression relative to the wild-type coding

sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Furthermore, the nucleic acid sequence can also be modified to introduce mutations into one or more regions of the gene, for instance to render the gene product non-functional and/or to eliminate the myristoylation site in Nef.

Synthetic expression cassettes exemplified herein include SEQ ID NO:49 and SEQ ID NO:97 (Env gp160-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:51 and SEQ ID NO:99 (Gag-encoding sequences modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:53 (Gag major homology region, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:55 and SEQ ID NO:101 (Nef-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:57 and SEQ ID NO:134 (Nef-encoding sequences with a mutation at position 125 resulting in a non-functional gene product, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA, respectively); SEQ ID NO:58 (RNAseH-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:60 (Integrase-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:62 and SEQ ID NO:103 (Pol-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:64 (Protease-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:66 (inactivated protease-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:68 (inactivated protease and RT mutated sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:70 (protease and reverse-transcriptase-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:72 and SEQ ID NO:105 (exon 1 of Rev, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:74 and SEQ ID NO:107 (exon 2 of Rev, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:76 (reverse transcriptase-encoding sequences, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:78 (mutated reverse-transcriptase, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:80 (exon 1 of Tat including a mutation that results in non-functional Tat, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:81 and SEQ ID NO:109 (exon 1 of Tat, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:83 and SEQ ID NO:111 (exon 2 of Tat, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:85 and SEQ ID NO:113) (Vif-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ

ID NO:87 and SEQ ID NO:115 (Vpr-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:89 and SEQ ID NO:117 (Vpu-encoding sequences, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA wild-type, respectively); SEQ ID NO:91 (sequences of exons 1 and 2 of Rev, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:93 (sequences of mutated exon 1 of Tat and exon 2 of Tat, where mutation of exon 1 results in non-functional Tat, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:94 (sequences of exons 1 and 2 of Tat, modified based on 8_5_TV1_C.ZA wild type); SEQ ID NO:96 and SEQ ID NO:135 (Nef-encoding sequences including a mutation to eliminate myristoylation site, modified based on 8_5_TV1_C.ZA wild type and 12-5_1_TV2_C.ZA, respectively).

2.2.1.5 Further Modification of Sequences Including HIV-1 Nucleic Acid Coding Sequences

The Type C HIV polypeptide-encoding expression cassettes described herein may also contain one or more further sequences encoding, for example, one or more transgenes. Further sequences (*e.g.*, transgenes) useful in the practice of the present invention include, but are not limited to, further sequences are those encoding further viral epitopes/antigens {including but not limited to, HCV antigens (e.g., E1, E2; Houghton, M.., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M.., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M.., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997), HIV antigens (e.g., derived from *tat, rev, nef* and/or *env*); and sequences encoding tumor antigens/epitopes. Further sequences may also be derived from non-viral sources, for instance, sequences encoding cytokines such interleukin-2 (IL-2), stem cell factor (SCF), interleukin 3 (IL-3), interleukin 6 (IL-6), interleukin 12 (IL-12), G-CSF, granulocyte macrophage-colony stimulating factor (GM-CSF), interleukin-1 alpha (IL-1I), interleukin-11 (IL-11), MIP-1I, tumor necrosis factor (TNF), leukemia inhibitory factor (LIF), c-kit ligand, thrombopoietin (TPO) and flt3 ligand, commercially available from several vendors such as, for example, Genzyme (Framingham, MA), Genentech (South San Francisco, CA), Amgen (Thousand Oaks, CA), R&D Systems and Immunex (Seattle, WA). Additional sequences are described below, for example in Section 2.3. Also, variations on the orientation of the Gag and other coding sequences, relative to each other, are described below.

HIV polypeptide coding sequences can be obtained from other Type C HIV isolates, see, e.g., Myers et al. Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids, 1997*, Los Alamos, New Mexico: Los Alamos National Laboratory. Synthetic expression cassettes can be generated using such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1).

Further, the synthetic expression cassettes of the present invention include related polypeptide sequences having greater than 85%, preferably greater than 90%, more preferably greater than 95%, and most preferably greater than 98% sequence identity to the synthetic expression cassette sequences disclosed herein (for example, (SEQ ID NOs:30-32; SEQ ID NOs: 3, 4, 20, and 21 and SEQ ID NOs:5-17). Various coding regions are indicated in Figures 3 and 4, for example in Figure 3 (AF110968), nucleotides 1-81 (SEQ ID NO:18); nucleotides 82-1512 (SEQ ID NO:6) encode a gp120 polypeptide, nucleotides 1513 to 2547 (SEQ ID NO:10) encode a gp41 polypeptide, nucleotides 82-2025 (SEQ ID NO:7) encode a gp140 polypeptide and nucleotides 82-2547 (SEQ ID NO:8) encode a gp160 polypeptide. Similarly, in Figure 98 (SEQ ID NO:127, strain 8_2_TV1_C.ZA), nucleotides 1-6 are an EcoR1 restriction site; nucleotides 7-87 a encode a wild-type (from 8_2_TV1_C.ZA) leader signal peptide; nucleotides 88 to 1563 encode a gp120 polypeptide; nucleotides 88 to 2064 encode a gp140 polypeptide; nucleotides 88 to 2607 encode a gp160 polypeptide.

2.2.3 Expression of Synthetic Sequences Encoding HIV-1 Subtype C and Related Polypeptides

Synthetic HIV-encoding sequences (expression cassettes) of the present invention can be cloned into a number of different expression vectors to evaluate levels of expression and, in the case of Gag, production of VLPs. The synthetic DNA fragments for HIV polypeptides can be cloned into eucaryotic expression vectors, including, a transient expression vector, CMV-promoter-based mammalian vectors, and a shuttle vector for use in baculovirus expression systems. Corresponding wild-type sequences can also be cloned into the same vectors.

These vectors can then be transfected into a several different cell types, including a variety of mammalian cell lines (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines are then cultured under appropriate conditions and the levels of any appropriate polypeptide product can be evaluated in supernatants. (see, Table A and Example 2). For example, p24 can be used to evaluate Gag expression; gp160, gp140 or gp120 can be used to evaluate Env expression; p6pol can be used to evaluate Pol

expression; prot can be used to evaluate protease; p15 for RNAseH; p31 for Integrase; and other appropriate polypeptides for Vif, Vpr, Tat, Rev, Vpu and Nef. Further, modified polypeptides can also be used, for example, other Env polypeptides include, but are not limited to, for example, native gp160, oligomeric gp140, monomeric gp120 as well as modified and/or synthetic sequences of these polypeptides. The results of these assays demonstrate that expression of synthetic HIV polypeptide-encoding sequences are significantly higher than corresponding wild-type sequences.

Further, Western Blot analysis can be used to show that cells containing the synthetic expression cassette produce the expected protein at higher per-cell concentrations than cells containing the native expression cassette. The HIV proteins can be seen in both cell lysates and supernatants. The levels of production are significantly higher in cell supernatants for cells transfected with the synthetic expression cassettes of the present invention.

Fractionation of the supernatants from mammalian cells transfected with the synthetic expression cassette can be used to show that the cassettes provide superior production of HIV proteins and, in the case of Gag, VLPs, relative to the wild-type sequences.

Efficient expression of these HIV-containing polypeptides in mammalian cell lines provides the following benefits: the polypeptides are free of baculovirus contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Subtype C HIV-containing polypeptides in CHO cells which is not feasible in the absence of the increased expression obtained using the constructs of the present invention. Exemplary Mammalian cell lines include, but are not limited to, BHK, VERO, HT1080, 293, 293T, RD, COS-7, CHO, Jurkat, HUT, SUPT, C8166, MOLT4/clone8, MT-2, MT-4, H9, PM1, CEM, and CEMX174, such cell lines are available, for example, from the A.T.C.C.).

A synthetic Gag expression cassette of the present invention will also exhibit high levels of expression and VLP production when transfected into insect cells. Synthetic expression cassettes described herein also demonstrate high levels of expression in insect cells. Further, in addition to a higher total protein yield, the final product from the synthetic polypeptides consistently contains lower amounts of contaminating baculovirus proteins than the final product from the native Type C sequences.

Further, synthetic expression cassettes of the present invention can also be introduced into yeast vectors which, in turn, can be transformed into and efficiently expressed by yeast cells (*Saccharomyces cerevisea*; using vectors as described in Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998).

In addition to the mammalian and insect vectors, the synthetic expression cassettes of the present invention can be incorporated into a variety of expression vectors using selected expression control elements. Appropriate vectors and control elements for any given cell type can be selected by one having ordinary skill in the art in view of the teachings of the present specification and information known in the art about expression vectors.

For example, a synthetic expression cassette can be inserted into a vector which includes control elements operably linked to the desired coding sequence, which allow for the expression of the gene in a selected cell-type. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter (a CMV promoter can include intron A), RSV, HIV-Ltr, the mouse mammary tumor virus LTR promoter (MMLV-ltr), the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook, et al., *supra*, as well as a bovine growth hormone terminator sequence. Introns, containing splice donor and acceptor sites, may also be designed into the constructs for use with the present invention (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

The desired synthetic polypeptide encoding sequences can be cloned into any number of commercially available vectors to generate expression of the polypeptide in an appropriate host system. These systems include, but are not limited to, the following: baculovirus expression {Reilly, P.R., *et al.*, BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992); Beames, *et al.*, *Biotechniques* 11:378 (1991); Pharmingen; Clontech, Palo Alto, CA)}, vaccinia expression {Earl, P. L., *et al.*, "Expression of proteins in mammalian cells using vaccinia" In *Current Protocols in Molecular Biology* (F. M. Ausubel, *et al.* Eds.),

Greene Publishing Associates & Wiley Interscience, New York (1991); Moss, B., *et al.*, U.S. Patent Number 5,135,855, issued 4 August 1992}, expression in bacteria {Ausubel, F.M., *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA; Clontech}, expression in yeast {Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998; Shuster, J.R., U.S. Patent No. 5,629,203, issued May 13, 1997; Gellissen, G., *et al.*, *Antonie Van Leeuwenhoek*, 62(1-2):79-93 (1992); Romanos, M.A., *et al.*, *Yeast* 8(6):423-488 (1992); Goeddel, D.V., *Methods in Enzymology* 185 (1990); Guthrie, C., and G.R. Fink, *Methods in Enzymology* 194 (1991)}, expression in mammalian cells {Clontech; Gibco-BRL, Ground Island, NY; *e.g.*, Chinese hamster ovary (CHO) cell lines (Haynes, J., *et al.*, *Nuc. Acid. Res.* 11:687-706 (1983); 1983, Lau, Y.F., *et al.*, *Mol. Cell. Biol.* 4:1469-1475 (1984); Kaufman, R. J., "Selection and coamplification of heterologous genes in mammalian cells," in *Methods in Enzymology*, vol. 185, pp537-566. Academic Press, Inc., San Diego CA (1991)}, and expression in plant cells {plant cloning vectors, Clontech Laboratories, Inc., Palo Alto, CA, and Pharmacia LKB Biotechnology, Inc., Pistcataway, NJ; Hood, E., *et al.*, *J. Bacteriol.* 168:1291-1301 (1986); Nagel, R., *et al.*, *FEMS Microbiol. Lett.* 67:325 (1990); An, *et al.*, "Binary Vectors", and others in Plant Molecular Biology Manual A3:1-19 (1988); Miki, B.L.A., *et al.*, pp.249-265, and others in Plant DNA Infectious Agents (Hohn, T., *et al.*, eds.) Springer-Verlag, Wien, Austria, (1987); *Plant Molecular Biology: Essential Techniques*, P.G. Jones and J.M. Sutton, New York, J. Wiley, 1997; Miglani, Gurbachan *Dictionary of Plant Genetics and Molecular Biology*, New York, Food Products Press, 1998; Henry, R. J., *Practical Applications of Plant Molecular Biology*, New York, Chapman & Hall, 1997}.

Also included in the invention is an expression vector, containing coding sequences and expression control elements which allow expression of the coding regions in a suitable host. The control elements generally include a promoter, translation initiation codon, and translation and transcription termination sequences, and an insertion site for introducing the insert into the vector. Translational control elements have been reviewed by M. Kozak (e.g., Kozak, M., *Mamm. Genome* **7(8)**:563-574, 1996; Kozak, M., *Biochimie* **76(9)**:815-821, 1994; Kozak, M., *J Cell Biol* **108(2)**:229-241, 1989; Kozak, M., and Shatkin, A.J., *Methods Enzymol* **60**:360-375, 1979).

Expression in yeast systems has the advantage of commercial production. Recombinant protein production by vaccinia and CHO cell line have the advantage of being mammalian expression systems. Further, vaccinia virus expression has several advantages including the following: (i) its wide host range; (ii) faithful post-transcriptional modification, processing, folding, transport, secretion, and assembly of recombinant proteins; (iii) high

level expression of relatively soluble recombinant proteins; and (iv) a large capacity to accommodate foreign DNA.

The recombinantly expressed polypeptides from synthetic HIV polypeptide-encoding expression cassettes are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including salt fractionation, ion exchange chromatography, gel filtration, size-exclusion chromatography, size-fractionation, and affinity chromatography. Immunoaffinity chromatography can be employed using antibodies generated based on, for example, HIV antigens.

Advantages of expressing the proteins of the present invention using mammalian cells include, but are not limited to, the following: well-established protocols for scale-up production; the ability to produce VLPs; cell lines are suitable to meet good manufacturing process (GMP) standards; culture conditions for mammalian cells are known in the art.

Various forms of the different embodiments of the invention, described herein, may be combined.

2.3 Production of Virus-like Particles and Use of the Constructs of the Present Invention to create Packaging Cell Lines.

The group-specific antigens (Gag) of human immunodeficiency virus type-1 (HIV-1) self-assemble into noninfectious virus-like particles (VLP) that are released from various eucaryotic cells by budding (reviewed by Freed, E.O., *Virology* **251**:1-15, 1998). The synthetic expression cassettes of the present invention provide efficient means for the production of HIV-Gag virus-like particles (VLPs) using a variety of different cell types, including, but not limited to, mammalian cells.

Viral particles can be used as a matrix for the proper presentation of an antigen entrapped or associated therewith to the immune system of the host.

2.3.1 VLP Production using the synthetic expression cassettes of the present invention

Experiments can be performed in support of the present invention to demonstrate that the synthetic expression cassettes of the present invention provide superior production of both Gag proteins and VLPs, relative to native Gag coding sequences. Further, electron microscopic evaluation of VLP production can show that free and budding immature virus particles of the expected size are produced by cells containing the synthetic expression cassettes.

Using the synthetic expression cassettes of the present invention, rather than native Gag coding sequences, for the production of virus-like particles provide several advantages. First, VLPs can be produced in enhanced quantity making isolation and purification of the VLPs easier. Second, VLPs can be produced in a variety of cell types using the synthetic expression cassettes, in particular, mammalian cell lines can be used for VLP production, for example, CHO cells. Production using CHO cells provides (i) VLP formation; (ii) correct myristoylation and budding; (iii) absence of non-mamallian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification. The synthetic expression cassettes of the present invention are also useful for enhanced expression in cell-types other than mammalian cell lines. For example, infection of insect cells with baculovirus vectors encoding the synthetic expression cassettes results in higher levels of total Gag protein yield and higher levels of VLP production (relative to wild-type coding sequences). Further, the final product from insect cells infected with the baculovirus-Gag synthetic expression cassettes consistently contains lower amounts of contaminating insect proteins than the final product when wild-type coding sequences are used.

VLPs can spontaneously form when the particle-forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs produced using the synthetic expression cassettes of the present invention are conveniently prepared using recombinant techniques. As discussed below, the Gag polypeptide encoding synthetic expression cassettes of the present invention can include other polypeptide coding sequences of interest (for example, HIV protease, HIV polymerase, HCV core; Env; synthetic Env; see, Example 1). Expression of such synthetic expression cassettes yields VLPs comprising the Gag polypeptide, as well as, the polypeptide of interest.

Once coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Sambrook et al, *supra*. The vector is then used to transform an appropriate host cell. Suitable recombinant expression systems include, but are not limited to, bacterial, mammalian, baculovirus/insect, vaccinia, Semliki Forest virus (SFV), Alphaviruses (such as, Sindbis, Venezuelan Equine Encephalitis (VEE)), mammalian, yeast and Xenopus expression systems, well known in the art. Particularly preferred expression systems are mammalian cell lines, vaccinia, Sindbis, insect and yeast systems.

For example, a number of mammalian cell lines are known in the art and include immortalized cell lines available from the American Type Culture Collection (A.T.C.C.), such as, but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillerimondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. See, e.g., Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987).

Viral vectors can be used for the production of particles in eucaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additionally, a vaccinia based infection/transfection system, as described in Tomei et al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. Alternately, T7 can be added as a purified protein or enzyme as in the "Progenitor" system (Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130). The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

Depending on the expression system and host selected, the VLPS are produced by growing host cells transformed by an expression vector under conditions whereby the particle-forming polypeptide is expressed and VLPs can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPs are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPs substantially intact. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated (or substantially purified) using methods that preserve the integrity thereof, such as, by gradient centrifugation, e.g., cesium chloride (CsCl) sucrose gradients, pelleting and the like (see, e.g., Kirnbauer et al. *J. Virol.* (1993) 67:6929-6936), as well as standard purification techniques including, e.g., ion exchange and gel filtration chromatography.

VLPs produced by cells containing the synthetic expression cassettes of the present invention can be used to elicit an immune response when administered to a subject. One advantage of the present invention is that VLPs can be produced by mammalian cells carrying the synthetic expression cassettes at levels previously not possible. As discussed above, the VLPs can comprise a variety of antigens in addition to the Gag polypeptide (e.g., Gag-protease, Gag-polymerase, Env, synthetic Env, etc.). Purified VLPs, produced using the synthetic expression cassettes of the present invention, can be administered to a vertebrate subject, usually in the form of vaccine compositions. Combination vaccines may also be used, where such vaccines contain, for example, an adjuvant subunit protein (e.g., Env). Administration can take place using the VLPs formulated alone or formulated with other antigens. Further, the VLPs can be administered prior to, concurrent with, or subsequent to, delivery of the synthetic expression cassettes for DNA immunization (see below) and/or delivery of other vaccines. Also, the site of VLP administration may be the same or different as other vaccine compositions that are being administered. Gene delivery can be accomplished by a number of methods including, but are not limited to, immunization with DNA, alphavirus vectors, pox virus vectors, and vaccinia virus vectors.

VLP immune-stimulating (or vaccine) compositions can include various excipients, adjuvants, carriers, auxiliary substances, modulating agents, and the like. The immune stimulating compositions will include an amount of the VLP/antigen sufficient to mount an immunological response. An appropriate effective amount can be determined by one of skill in the art. Such an amount will fall in a relatively broad range that can be determined through routine trials and will generally be an amount on the order of about 0.1 μg to about 1000 μg, more preferably about 1 μg to about 300 μg, of VLP/antigen.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycollic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-

glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* **14**(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* **11**(2):149-54, 1993. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc., as well as toxins derived from *E. coli.*

Adjuvants may also be used to enhance the effectiveness of the compositions. Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freunds Adjuvant (CFA) and Incomplete Freunds Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (6) oligonucleotides or polymeric molecules encoding immunostimulatory CpG mofifs (Davis, H.L., et al., *J. Immunology* **160**:870-876, 1998; Sato, Y. et al., *Science* **273**:352-354, 1996) or complexes of antigens/oligonucleotides {Polymeric molecules include double and single stranded RNA and DNA, and backbone modifications thereof, for example, methylphosphonate linkages; or (7) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63) LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino acid at position 9 and glycine substituted

at position 129) (see, e.g., International Publication Nos. W093/13202 and W092/19265); and (8) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Further, such polymeric molecules include alternative polymer backbone structures such as, but not limited to, polyvinyl backbones (Pitha, *Biochem Biophys Acta*, 204:39, 1970a; Pitha, *Biopolymers*, 9:965, 1970b), and morpholino backbones (Summerton, J., *et al.*, U.S. Patent No. 5,142,047, issued 08/25/92; Summerton, J., *et al.*, U.S. Patent No. 5,185,444 issued 02/09/93). A variety of other charged and uncharged polynucleotide analogs have been reported. Numerous backbone modifications are known in the art, including, but not limited to, uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, and carbamates) and charged linkages (*e.g.*, phosphorothioates and phosphorodithioates).}; and (7) other substances that act as immunostimulating agents to enhance the effectiveness of the VLP immune-stimulating (or vaccine) composition. Alum, CpG oligonucleotides, and MF59 are preferred.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acteyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(l'-2'-dipalmitoyl-*sn*-glycero-3-huydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Dosage treatment with the VLP composition may be a single dose schedule or a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals, chosen to maintain and/or reinforce the immune response, for example at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the subject and be dependent on the judgment of the practitioner.

If prevention of disease is desired, the antigen carrying VLPs are generally administered prior to primary infection with the pathogen of interest. If treatment is desired, e.g., the reduction of symptoms or recurrences, the VLP compositions are generally administered subsequent to primary infection.

2.3.2 USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

A number of viral based systems have been developed for use as gene transfer vectors for mammalian host cells. For example, retroviruses (in particular, lentiviral vectors) provide a convenient platform for gene delivery systems. A coding sequence of interest (for example, a sequence useful for gene therapy applications) can be inserted into a gene delivery vector

and packaged in retroviral particles using techniques known in the art. Recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described, including, for example, the following: (U.S. Patent No. 5,219,740; Miller et al. (1989) *BioTechniques* 7:980; Miller, A.D. (1990) *Human Gene Therapy* 1:5; Scarpa et al. (1991) *Virology* 180:849; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033; Boris-Lawrie et al. (1993) *Cur. Opin. Genet. Develop.* 3:102; GB 2200651; EP 0415731; EP 0345242; WO 89/02468; WO 89/05349; WO 89/09271; WO 90/02806; WO 90/07936; WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; WO 93/11230; WO 93/10218; WO 91/02805; in U.S. 5,219,740; U.S. 4,405,712; U.S. 4,861,719; U.S. 4,980,289 and U.S. 4,777,127; in U.S. Serial No. 07/800,921; and in Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53:83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci USA* 81;6349; and Miller (1990) *Human Gene Therapy* 1.

In other embodiments, gene transfer vectors can be constructed to encode a cytokine or other immunomodulatory molecule. For example, nucleic acid sequences encoding native IL-2 and gamma-interferon can be obtained as described in US Patent Nos. 4,738,927 and 5,326,859, respectively, while useful muteins of these proteins can be obtained as described in U.S. Patent No. 4,853,332. Nucleic acid sequences encoding the short and long forms of mCSF can be obtained as described in US Patent Nos. 4,847,201 and 4,879,227, respectively. In particular aspects of the invention, retroviral vectors expressing cytokine or immunomodulatory genes can be produced as described herein (for example, employing the packaging cell lines of the present invention) and in International Application No. PCT US 94/02951, entitled "Compositions and Methods for Cancer Immunotherapy."

Examples of suitable immunomodulatory molecules for use herein include the following: IL-1 and IL-2 (Karupiah et al. (1990) *J. Immunology* 144:290-298, Weber et al. (1987) *J. Exp. Med.* 166:1716-1733, Gansbacher et al. (1990) *J. Exp. Med.* 172:1217-1224, and U.S. Patent No. 4,738,927); IL-3 and IL-4 (Tepper et al. (1989) *Cell* 57:503-512, Golumbek et al. (1991) *Science* 254:713-716, and U.S. Patent No. 5,017,691); IL-5 and IL-6 (Brakenhof et al. (1987) *J. Immunol.* 139:4116-4121, and International Publication No. WO 90/06370); IL-7 (U.S. Patent No. 4,965,195); IL-8, IL-9, IL-10, IL-11, IL-12, and IL-13 (*Cytokine Bulletin*, Summer 1994); IL-14 and IL-15; alpha interferon (Finter et al. (1991) *Drugs* 42:749-765, U.S. Patent Nos. 4,892,743 and 4,966,843, International Publication No. WO 85/02862, Nagata et al. (1980) *Nature* 284:316-320, Familletti et al. (1981) *Methods in Enz.* 78:387-394, Twu et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2046-2050, and Faktor et

al. (1990) *Oncogene* 5:867-872); beta-interferon (Seif et al. (1991) *J. Virol.* 65:664-671); gamma-interferons (Radford et al. (1991) *The American Society of Hepatology* 20082015, Watanabe et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:9456-9460, Gansbacher et al. (1990) *Cancer Research* 50:7820-7825, Maio et al. (1989) *Can. Immunol. Immunother.* 30:34-42, and U.S. Patent Nos. 4,762,791 and 4,727,138); G-CSF (U.S. Patent Nos. 4,999,291 and 4,810,643); GM-CSF (International Publication No. WO 85/04188).

Immunomodulatory factors may also be agonists, antagonists, or ligands for these molecules. For example, soluble forms of receptors can often behave as antagonists for these types of factors, as can mutated forms of the factors themselves.

Nucleic acid molecules that encode the above-described substances, as well as other nucleic acid molecules that are advantageous for use within the present invention, may be readily obtained from a variety of sources, including, for example, depositories such as the American Type Culture Collection, or from commercial sources such as British Bio-Technology Limited (Cowley, Oxford England). Representative examples include BBG 12 (containing the GM-CSF gene coding for the mature protein of 127 amino acids), BBG 6 (which contains sequences encoding gamma interferon), A.T.C.C. Deposit No. 39656 (which contains sequences encoding TNF), A.T.C.C. Deposit No. 20663 (which contains sequences encoding alpha-interferon), A.T.C.C. Deposit Nos. 31902, 31902 and 39517 (which contain sequences encoding beta-interferon), A.T.C.C. Deposit No. 67024 (which contains a sequence which encodes Interleukin-1b), A.T.C.C. Deposit Nos. 39405, 39452, 39516, 39626 and 39673 (which contain sequences encoding Interleukin-2), A.T.C.C. Deposit Nos. 59399, 59398, and 67326 (which contain sequences encoding Interleukin-3), A.T.C.C. Deposit No. 57592 (which contains sequences encoding Interleukin-4), A.T.C.C. Deposit Nos. 59394 and 59395 (which contain sequences encoding Interleukin-5), and A.T.C.C. Deposit No. 67153 (which contains sequences encoding Interleukin-6).

Plasmids containing cytokine genes or immunomodulatory genes (International Publication Nos. WO 94/02951 and WO 96/21015) can be digested with appropriate restriction enzymes, and DNA fragments containing the particular gene of interest can be inserted into a gene transfer vector using standard molecular biology techniques. (*See, e.g.*, Sambrook et al., *supra.*, or Ausbel et al. (eds) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience).

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. For example, plasmids which contain sequences that encode altered cellular products may be

obtained from a depository such as the A.T.C.C., or from commercial sources. Plasmids containing the nucleotide sequences of interest can be digested with appropriate restriction enzymes, and DNA fragments containing the nucleotide sequences can be inserted into a gene transfer vector using standard molecular biology techniques.

Alternatively, cDNA sequences for use with the present invention may be obtained from cells which express or contain the sequences, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA. Briefly, mRNA from a cell which expresses the gene of interest can be reverse transcribed with reverse transcriptase using oligo-dT or random primers. The single stranded cDNA may then be amplified by PCR (see U.S. Patent Nos. 4,683,202, 4,683,195 and 4,800,159, see also *PCR Technology: Principles and Applications for DNA Amplification*, Erlich (ed.), Stockton Press, 1989)) using oligonucleotide primers complementary to sequences on either side of desired sequences.

The nucleotide sequence of interest can also be produced synthetically, rather than cloned, using a DNA synthesizer (*e.g.*, an Applied Biosystems Model 392 DNA Synthesizer, available from ABI, Foster City, California). The nucleotide sequence can be designed with the appropriate codons for the expression product desired. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge (1981) *Nature* 292:756; Nambair et al. (1984) *Science* 223:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311.

The synthetic expression cassettes of the present invention can be employed in the construction of packaging cell lines for use with retroviral vectors.

One type of retrovirus, the murine leukemia virus, or "MLV", has been widely utilized for gene therapy applications (see generally Mann et al. *(Cell 33:153,* 1993), Cane and Mulligan *(Proc, Nat'l. Acad. Sci. USA* 81:6349, 1984), and Miller et al., *Human Gene 2lerapy* 1:5-14,1990.

Lentiviral vectors typically, comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, a promoter operably linked to one or more genes of interest, an origin of second strand DNA synthesis and a 3' lentiviral LTR, wherein the lentiviral vector contains a nuclear transport element. The nuclear transport element may be located either upstream (5') or downstream (3') of a coding sequence of interest (for example, a synthetic Gag or Env expression cassette of the present invention). Within certain embodiments, the nuclear transport element is not RRE. Within one embodiment the packaging signal is an extended packaging signal. Within other embodiments the promoter is a tissue specific promoter, or,

alternatively, a promoter such as CMV. Within other embodiments, the lentiviral vector further comprises an internal ribosome entry site.

A wide variety of lentiviruses may be utilized within the context of the present invention, including for example, lentiviruses selected from the group consisting of HIV, HIV-1, HIV-2, FIV and SIV.

In one embodiment of the present invention synthetic Gag-polymerase expression cassettes are provided comprising a promoter and a sequence encoding synthetic Gag-polymerase and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif.

Within yet another aspect of the invention, host cells (e.g., packaging cell lines) are provided which contain any of the expression cassettes described herein. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Gag-polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Gag-polymerase. Packaging cell lines may further comprise a promoter and a sequence encoding tat, rev, or an envelope, wherein the promoter is operably linked to the sequence encoding tat, rev, Env or sequences encoding modified versions of these proteins. The packaging cell line may further comprise a sequence encoding any one or more of nef, vif, vpu or vpr (wild-type or synthetic).

In one embodiment, the expression cassette (carrying, for example, the synthetic Gag-polymerase) is stably integrated. The packaging cell line, upon introduction of a lentiviral vector, typically produces particles. The promoter regulating expression of the synthetic expression cassette may be inducible. Typically, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are essentially free of replication competent virus.

Packaging cell lines are provided comprising an expression cassette which directs the expression of a synthetic *Gag-polymerase* gene or comprising an expression cassette which directs the expression of a synthetic Env genes described herein. (See, also, Andre, S., et al., *Journal of Virology* **72(2)**:1497-1503, 1998; Haas, J., et al., *Current Biology* **6(3)**:315-324, 1996) for a description of other modified Env sequences). A lentiviral vector is introduced into the packaging cell line to produce a vector producing cell line.

As noted above, lentiviral vectors can be designed to carry or express a selected gene(s) or sequences of interest. Lentiviral vectors may be readily constructed from a wide variety of lentiviruses *(see* RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985). Representative examples of lentiviruses included HIV, HIV-1, HIV-2, FIV and SIV. Such lentiviruses may either be obtained from patient isolates, or, more

preferably, from depositories or collections such as the American Type Culture Collection, or isolated from known sources using available techniques.

Portions of the lentiviral gene delivery vectors (or vehicles) may be derived from different viruses. For example, in a given recombinant lentiviral vector, LTRs may be derived from an HIV, a packaging signal from SIV, and an origin of second strand synthesis from HrV-2. Lentiviral vector constructs may comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, one or more heterologous sequences, an origin of second strand DNA synthesis and a 3' LTR, wherein said lentiviral vector contains a nuclear transport element that is not RRE.

Briefly, Long Terminal Repeats ("LTRs") are subdivided into three elements, designated U5, R and U3. These elements contain a variety of signals which are responsible for the biological activity of a retrovirus, including for example, promoter and enhancer elements which are located within U3. LTRs may be readily identified in the provirus (integrated DNA form) due to their precise duplication at either end of the genome. As utilized herein, a 5' LTR should be understood to include a 5' promoter element and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector. The 3' LTR should be understood to include a polyadenylation signal, and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector.

The tRNA binding site and origin of second strand DNA synthesis are also important for a retrovirus to be biologically active, and may be readily identified by one of skill in the art. For example, retroviral tRNA binds to a tRNA binding site by Watson-Crick base pairing, and is carried with the retrovirus genome into a viral particle. The tRNA is then utilized as a primer for DNA synthesis by reverse transcriptase. The tRNA binding site may be readily identified based upon its location just downstream from the 5'LTR. Similarly, the origin of second strand DNA synthesis is, as its name implies, important for the second strand DNA synthesis of a retrovirus. This region, which is also referred to as the poly-purine tract, is located just upstream of the 3'LTR.

In addition to a 5' and 3' LTR, tRNA binding site, and origin of second strand DNA synthesis, recombinant retroviral vector constructs may also comprise a packaging signal, as well as one or more genes or coding sequences of interest. In addition, the lentiviral vectors have a nuclear transport element which, in preferred embodiments is not RRE. Representative examples of suitable nuclear transport elements include the element in Rous sarcoma virus (Ogert, et al., J *ViroL 70,* 3834-3843, 1996), the element in Rous sarcoma virus (Liu & *Mertz, Genes & Dev., 9,* 1766-1789, 1995) and the element in the genome of simian retrovirus type I (Zolotukhin, et al., *J Virol. 68,* 7944-7952, 1994). Other potential elements

include the elements in the histone gene (Kedes, *Annu. Rev. Biochem. 48,* 837-870, 1970), the α-interferon gene (Nagata et al., *Nature 287,* 401-408, 1980), the β-adrenergic receptor gene (Koilka, et al., *Nature 329,* 75-79, 1987), and the c-Jun gene (Hattorie, et al., *Proc. Natl. Acad. Sci. USA 85,* 9148-9152, 1988).

Recombinant lentiviral vector constructs typically lack both *Gag-polymerase* and *Env* coding sequences. Recombinant lentiviral vector typically contain less than 20, preferably 15, more preferably 10, and most preferably 8 consecutive nucleotides found in *Gag-polymerase* and *Env* genes. One advantage of the present invention is that the synthetic *Gag-polymerase* expression cassettes, which can be used to construct packaging cell lines for the recombinant retroviral vector constructs, have little homology to wild-type Gag-polymerase sequences and thus considerably reduce or eliminate the possibility of homologous recombination between the synthetic and wild-type sequences.

Lentiviral vectors may also include tissue-specific promoters to drive expression of one or more genes or sequences of interest.

Lentiviral vector constructs may be generated such that more than one gene of interest is expressed. This may be accomplished through the use of di- or oligo-cistronic cassettes (e.g., where the coding regions are separated by 80 nucleotides or less, *see generally* Levin et al., *Gene* 108:167-174, 1991), or through the use of Internal Ribosome Entry Sites ("IRES").

Packaging cell lines suitable for use with the above described recombinant retroviral vector constructs may be readily prepared given the disclosure provided herein. Briefly, the parent cell line from which the packaging cell line is derived can be selected from a variety of mammalian cell lines, including for example, 293, RD, COS-7, CHO, BHK, VERO, HT1080, and myeloma cells.

After selection of a suitable host cell for the generation of a packaging cell line, one or more expression cassettes are introduced into the cell line in order to complement or supply in *trans* components of the vector which have been deleted.

Representative examples of suitable expression cassettes have been described herein and include synthetic Env, synthetic Gag, synthetic Gag-protease, and synthetic Gag-polymerase expression cassettes, which comprise a promoter and a sequence encoding, e.g., Gag-polymerase and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif. As described above, the native and/or synthetic coding sequences may also be utilized in these expression cassettes.

Utilizing the above-described expression cassettes, a wide variety of packaging cell lines can be generated. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Gag-

polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Gag-polymerase. Within other aspects, packaging cell lines are provided comprising a promoter and a sequence encoding tat, rev, Env, or other HIV antigens or epitopes derived therefrom, wherein the promoter is operably linked to the sequence encoding tat, rev, Env, or the HIV antigen or epitope. Within further embodiments, the packaging cell line may comprise a sequence encoding any one or more of nef, vif, vpu or vpr. For example, the packaging cell line may contain only nef, vif, vpu, or vpr alone, nef and vif, nef and vpu, nef and vpr, vif and vpu, vif and vpr, vpu and vpr, nef vif and vpu, nef vif and vpr, nef vpu and vpr, vvir vpu and vpr, or, all four of nef, vif, vpu, and vpr.

In one embodiment, the expression cassette is stably integrated. Within another embodiment, the packaging cell line, upon introduction of a lentiviral vector, produces particles. Within further embodiments the promoter is inducible. Within certain preferred embodiments of the invention, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are free of replication competent virus.

The synthetic cassettes containing modified coding sequences are transfected into a selected cell line. Transfected cells are selected that (i) carry, typically, integrated, stable copies of the HIV coding sequences, and (ii) are expressing acceptable levels of these polypeptides (expression can be evaluated by methods known in the prior art, e.g., see Examples 1-4). The ability of the cell line to produce VLPs may also be verified.

A sequence of interest is constructed into a suitable viral vector as discussed above. This defective virus is then transfected into the packaging cell line. The packaging cell line provides the viral functions necessary for producing virus-like particles into which the defective viral genome, containing the sequence of interest, are packaged. These VLPs are then isolated and can be used, for example, in gene delivery or gene therapy.

Further, such packaging cell lines can also be used to produce VLPs alone, which can, for example, be used as adjuvants for administration with other antigens or in vaccine compositions. Also, co-expression of a selected sequence of interest encoding a polypeptide (for example, an antigen) in the packaging cell line can also result in the entrapment and/or association of the selected polypeptide in/with the VLPs.

Various forms of the different embodiments of the present invention (*e.g.*, constructs) may be combined.

2.4 DNA IMMUNIZATION AND GENE DELIVERY

A variety of HIV polypeptide antigens, particularly Type C HIV antigens, can be used in the practice of the present invention. HIV antigens can be included in DNA immunization

constructs containing, for example, a synthetic Gag expression cassette fused in-frame to a coding sequence for the polypeptide antigen (synthetic or wild-type), where expression of the construct results in VLPs presenting the antigen of interest.

HIV antigens of particular interest to be used in the practice of the present invention include tat, rev, nef, vif, vpu, vpr, and other HIV antigens or epitopes derived therefrom. These antigens may be synthetic (as described herein) or wild-type. Further, the packaging cell line may contain only nef, and HIV-1 (also known as HTLV-III, LAV, ARV, etc.), including, but not limited to, antigens such as gp120, gp41, gp160 (both native and modified); Gag; and pol from a variety of isolates including, but not limited to, HIV_{IIIb}, HIV_{SF2}, $HIV\text{-}1_{SF162}$, $HIV\text{-}1_{SF170}$, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, $HIV\text{-}1_{CM235}$,, $HIV\text{-}1_{US4}$, other HIV-1 strains from diverse subtypes(e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., $HIV\text{-}2_{UC1}$ and $HIV\text{-}2_{UC2}$). See, e.g., Myers, et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico; Myers, et al., *Human Retroviruses and Aids, 1990*, Los Alamos, New Mexico: Los Alamos National Laboratory.

To evaluate efficacy, DNA immunization using synthetic expression cassettes of the present invention can be performed, for instance as described in Example 4. Mice are immunized with both the Gag (and/or Env) synthetic expression cassette and the Gag (and/or Env) wild type expression cassette. Mouse immunizations with plasmid-DNAs will show that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the second boost immunization will induce a secondary immune response, for example, after approximately two weeks. Further, the results of CTL assays will show increased potency of synthetic Gag (and/or Env) expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

It is readily apparent that the subject invention can be used to mount an immune response to a wide variety of antigens and hence to treat or prevent a HIV infection, particularly Type C HIV infection.

2.4.1 DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. Furthermore, the desired gene can be isolated directly from cells and tissues containing the same, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and

isolate DNA. The gene of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature* (1981) 292:756; Nambair et al., *Science* (1984) 223:1299; Jay et al., *J. Biol. Chem.* (1984) 259:6311; Stemmer, W.P.C., (1995) *Gene* **164**:49-53.

Next, the gene sequence encoding the desired antigen can be inserted into a vector containing a synthetic expression cassette of the present invention. In certain embodiments, the antigen is inserted into the synthetic Gag coding sequence such that when the combined sequence is expressed it results in the production of VLPs comprising the Gag polypeptide and the antigen of interest, *e.g.,* Env (native or modified) or other antigen(s) (native or modified) derived from HIV. Insertions can be made within the coding sequence or at either end of the coding sequence (5', amino terminus of the expressed Gag polypeptide; or 3', carboxy terminus of the expressed Gag polypeptide)(Wagner, R., et al., *Arch Virol.* **127**:117-137, 1992; Wagner, R., et al., *Virology* **200**:162-175, 1994; Wu, X., et al., *J. Virol.* **69(6)**:3389-3398, 1995; Wang, C-T., et al., *Virology* **200**:524-534, 1994; Chazal, N., et al., *Virology* **68(1)**:111-122, 1994; Griffiths, J.C., et al., *J. Virol.* **67(6)**:3191-3198, 1993; Reicin, A.S., et al., *J. Virol.* **69(2)**:642-650, 1995).

Up to 50% of the coding sequences of p55Gag can be deleted without affecting the assembly to virus-like particles and expression efficiency (Borsetti, A., et al, *J. Virol.* **72(11)**:9313-9317, 1998; Gamier, L., et al., *J Virol* **72(6)**:4667-4677, 1998; Zhang, Y., et al., *J Virol* **72(3)**:1782-1789, 1998; Wang, C., et al., *J Virol* **72(10)**: 7950-7959, 1998). In one embodiment of the present invention, immunogenicity of the high level expressing synthetic Gag expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted regions of Gag sequence. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency of the expression product.

When sequences are added to the amino terminal end of Gag, the polynucletide can contain coding sequences at the 5' end that encode a signal for addition of a myristic moiety to the Gag-containing polypeptide (e.g., sequences that encode Met-Gly).

The ability of Gag-containing polypeptide constructs to form VLPs can be empirically determined following the teachings of the present specification.

The synthetic expression cassettes can also include control elements operably linked to the coding sequence, which allow for the expression of the gene *in vivo* in the subject species. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook et al., *supra*, as well as a bovine growth hormone terminator sequence.

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence.

Furthermore, plasmids can be constructed which include a chimeric antigen-coding gene sequences, encoding, e.g., multiple antigens/epitopes of interest, for example derived from more than one viral isolate.

Typically the antigen coding sequences precede or follow the synthetic coding sequence and the chimeric transcription unit will have a single open reading frame encoding both the antigen of interest and the synthetic coding sequences. Alternatively, multi-cistronic cassettes (e.g., bi-cistronic cassettes) can be constructed allowing expression of multiple antigens from a single mRNA using the EMCV IRES, or the like.

Once complete, the constructs are used for nucleic acid immunization using standard gene delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery

systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109.

A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery. AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Patent Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769 (published 4 March 1993); Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., *Vaccines 90* (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzyczka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875.

Another vector system useful for delivering the polynucleotides of the present invention is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., et al. (U.S. Patent No. 5,676,950, issued October 14, 1997).

Additional viral vectors which will find use for delivering the nucleic acid molecules encoding the antigens of interest include those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the genes can be constructed as follows. The DNA encoding the particular synthetic HIV subtype C polypeptide coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the coding sequences of interest

into the viral genome. The resulting TK-recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the genes. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an avipox vector is particularly desirable in human and other mammalian species since members of the avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as, but not limited to, vectors derived from the Sindbis, Semliki Forest, and Venezuelan Equine Encephalitis viruses, will also find use as viral vectors for delivering the polynucleotides of the present invention (for example, a synthetic Gag-polypeptide encoding expression cassette). For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072; as well as, Dubensky, Jr., T.W., et al., U.S. Patent No. 5,843,723, issued December 1, 1998, and Dubensky, Jr., T.W., U.S. Patent No. 5,789,245, issued August 4, 1998.

A vaccinia based infection/transfection system can be conveniently used to provide for inducible, transient expression of the coding sequences of interest in a host cell. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or to the delivery of genes using other viral vectors, an amplification system can be used that

will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered. Translation of RNA derived from this template will generate T7 RNA polymerase which in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, e.g., International Publication No. WO 94/26911; Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao et al., *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen et al., *Nuc. Acids Res.* (1994) 22:2114-2120; and U.S. Patent No. 5,135,855.

Synthetic expression cassettes of interest can also be delivered without a viral vector. For example, the synthetic expression cassette can be packaged in liposomes prior to delivery to the subject or to cells derived therefrom. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416); mRNA (Malone et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6077-6081); and purified transcription factors (Debs et al., *J. Biol. Chem.* (1990) 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., *Proc. Natl.*

Acad. Sci. USA (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as, from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphoshatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammelar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in METHODS OF IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483; Wilson et al., *Cell* (1979) 17:77); Deamer and Bangham, *Biochim. Biophys. Acta* (1976) 443:629; Ostro et al., *Biochem. Biophys. Res. Commun.* (1977) 76:836; Fraley et al., *Proc. Natl. Acad. Sci. USA* (1979) 76:3348); Enoch and Strittmatter, *Proc. Natl. Acad. Sci. USA* (1979) 76:145); Fraley et al., *J. Biol. Chem.* (1980) 255:10431; Szoka and Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* (1978) 75:145; and Schaefer-Ridder et al., *Science* (1982) 215:166.

The DNA and/or protein antigen(s) can also be delivered in cochleate lipid compositions similar to those described by Papahadjopoulos et al., *Biochem. Biophys. Acta.* (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

The synthetic expression cassette of interest may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected antigen to the immune system and promote trapping and retention of antigens in local lymph nodes. The particles can be phagocytosed by macrophages and can enhance antigen presentation through cytokine release. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Suitable microparticles may also be manufactured in the presence of charged detergents, such as anionic or cationic detergents, to yield microparticles with a surface having a net negative or a net positive charge. For example, microparticles manufactured with anionic detergents, such as

hexadecyltrimethylammonium bromide (CTAB), i.e. CTAB-PLG microparticles, adsorb negatively charged macromolecules, such as DNA. (see, e.g., Int'l Application Number PCT/US99/17308).

Furthermore, other particulate systems and polymers can be used for the *in vivo* or *ex vivo* delivery of the gene of interest. For example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules, are useful for transferring a nucleic acid of interest. Similarly, DEAE dextran-mediated transfection, calcium phosphate precipitation or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like, will find use with the present methods. See, e.g., Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Peptoids (Zuckerman, R.N., et al., U.S. Patent No. 5,831,005, issued November 3, 1998) may also be used for delivery of a construct of the present invention.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering synthetic expression cassettes of the present invention. The particles are coated with the synthetic expression cassette(s) to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744. Also, needle-less injection systems can be used (Davis, H.L., et al, *Vaccine* 12:1503-1509, 1994; Bioject, Inc., Portland, OR).

Recombinant vectors carrying a synthetic expression cassette of the present invention are formulated into compositions for delivery to the vertebrate subject. These compositions may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). The compositions will comprise a "therapeutically effective amount" of the gene of interest such that an amount of the antigen can be produced *in vivo* so that an immune response is generated in the individual to which it is administered. The exact amount necessary will vary depending on the subject being treated; the age and general condition of the subject to be treated; the capacity of the subject's immune system to synthesize antibodies; the degree of protection desired; the severity of the condition being treated; the particular antigen selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. Thus, a "therapeutically effective amount" will fall in a relatively broad range that can be determined through routine trials.

The compositions will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, polyethyleneglycol, hyaluronic acid, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Certain facilitators of nucleic acid uptake and/or expression can also be included in the compositions or coadministered, such as, but not limited to, bupivacaine, cardiotoxin and sucrose.

Once formulated, the compositions of the invention can be administered directly to the subject (e.g., as described above) or, alternatively, delivered *ex vivo*, to cells derived from the subject, using methods such as those described above. For example, methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and can include, e.g., dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, lipofectamine and LT-1 mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) (with or without the corresponding antigen) in liposomes, and direct microinjection of the DNA into nuclei.

Direct delivery of synthetic expression cassette compositions *in vivo* will generally be accomplished with or without viral vectors, as described above, by injection using either a conventional syringe or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). The constructs can be injected either subcutaneously, epidermally, intradermally, intramucosally such as nasally, rectally and vaginally, intraperitoneally, intravenously, orally or intramuscularly. Delivery of DNA into cells of the epidermis is particularly preferred as this mode of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Other modes of administration include oral and pulmonary administration, suppositories, needle-less injection, transcutaneous and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. Administration of nucleic acids may also be combined with administration of peptides or other substances.

2.4.2 Ex vivo Delivery of the Synthetic Expression Cassettes of the Present Invention

In one embodiment, T cells, and related cell types (including but not limited to antigen presenting cells, such as, macrophage, monocytes, lymphoid cells, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), can be used for *ex vivo* delivery of the synthetic expression cassettes of the present invention. T cells can be isolated from peripheral blood lymphocytes (PBLs) by a variety of procedures known to those skilled in the art. For example, T cell populations can be "enriched" from a population of PBLs through

the removal of accessory and B cells. In particular, T cell enrichment can be accomplished by the elimination of non-T cells using anti-MHC class II monoclonal antibodies. Similarly, other antibodies can be used to deplete specific populations of non-T cells. For example, anti-Ig antibody molecules can be used to deplete B cells and anti-MacI antibody molecules can be used to deplete macrophages.

T cells can be further fractionated into a number of different subpopulations by techniques known to those skilled in the art. Two major subpopulations can be isolated based on their differential expression of the cell surface markers CD4 and CD8. For example, following the enrichment of T cells as described above, $CD4^+$ cells can be enriched using antibodies specific for CD4 (see Coligan et al., *supra*). The antibodies may be coupled to a solid support such as magnetic beads. Conversely, CD8+ cells can be enriched through the use of antibodies specific for CD4 (to remove $CD4^+$ cells), or can be isolated by the use of CD8 antibodies coupled to a solid support. CD4 lymphocytes from HIV-1 infected patients can be expanded *ex vivo*, before or after transduction as described by Wilson et. al. (1995) *J. Infect. Dis.* 172:88.

Following purification of T cells, a variety of methods of genetic modification known to those skilled in the art can be performed using non-viral or viral-based gene transfer vectors constructed as described herein. For example, one such approach involves transduction of the purified T cell population with vector-containing supernatant of cultures derived from vector producing cells. A second approach involves co-cultivation of an irradiated monolayer of vector-producing cells with the purified T cells. A third approach involves a similar co-cultivation approach; however, the purified T cells are pre-stimulated with various cytokines and cultured 48 hours prior to the co-cultivation with the irradiated vector producing cells. Pre-stimulation prior to such transduction increases effective gene transfer (Nolta et al. (1992) *Exp. Hematol.* 20:1065). Stimulation of these cultures to proliferate also provides increased cell populations for re-infusion into the patient. Subsequent to co-cultivation, T cells are collected from the vector producing cell monolayer, expanded, and frozen in liquid nitrogen.

Gene transfer vectors, containing one or more synthetic expression cassette of the present invention (associated with appropriate control elements for delivery to the isolated T cells) can be assembled using known methods.

Selectable markers can also be used in the construction of gene transfer vectors. For example, a marker can be used which imparts to a mammalian cell transduced with the gene transfer vector resistance to a cytotoxic agent. The cytotoxic agent can be, but is not limited to, neomycin, aminoglycoside, tetracycline, chloramphenicol, sulfonamide, actinomycin,

netropsin, distamycin A, anthracycline, or pyrazinamide. For example, neomycin phosphotransferase II imparts resistance to the neomycin analogue geneticin (G418).

The T cells can also be maintained in a medium containing at least one type of growth factor prior to being selected. A variety of growth factors are known in the art which sustain the growth of a particular cell type. Examples of such growth factors are cytokine mitogens such as rIL-2, IL-10, IL-12, and IL-15, which promote growth and activation of lymphocytes. Certain types of cells are stimulated by other growth factors such as hormones, including human chorionic gonadotropin (hCG) and human growth hormone. The selection of an appropriate growth factor for a particular cell population is readily accomplished by one of skill in the art.

For example, white blood cells such as differentiated progenitor and stem cells are stimulated by a variety of growth factors. More particularly, IL-3, IL-4, IL-5, IL-6, IL-9, GM-CSF, M-CSF, and G-CSF, produced by activated T_H and activated macrophages, stimulate myeloid stem cells, which then differentiate into pluripotent stem cells, granulocyte-monocyte progenitors, eosinophil progenitors, basophil progenitors, megakaryocytes, and erythroid progenitors. Differentiation is modulated by growth factors such as GM-CSF, IL-3, IL-6, IL-11, and EPO.

Pluripotent stem cells then differentiate into lymphoid stem cells, bone marrow stromal cells, T cell progenitors, B cell progenitors, thymocytes, T_H Cells, T_C cells, and B cells. This differentiation is modulated by growth factors such as IL-3, IL-4, IL-6, IL-7, GM-CSF, M-CSF, G-CSF, IL-2, and IL-5.

Granulocyte-monocyte progenitors differentiate to monocytes, macrophages, and neutrophils. Such differentiation is modulated by the growth factors GM-CSF, M-CSF, and IL-8. Eosinophil progenitors differentiate into eosinophils. This process is modulated by GM-CSF and IL-5.

The differentiation of basophil progenitors into mast cells and basophils is modulated by GM-CSF, IL-4, and IL-9. Megakaryocytes produce platelets in response to GM-CSF, EPO, and IL-6. Erythroid progenitor cells differentiate into red blood cells in response to EPO.

Thus, during activation by the CD3-binding agent, T cells can also be contacted with a mitogen, for example a cytokine such as IL-2. In particularly preferred embodiments, the IL-2 is added to the population of T cells at a concentration of about 50 to 100 μg/ml. Activation with the CD3-binding agent can be carried out for 2 to 4 days.

Once suitably activated, the T cells are genetically modified by contacting the same with a suitable gene transfer vector under conditions that allow for transfection of the vectors

into the T cells. Genetic modification is carried out when the cell density of the T cell population is between about 0.1 x 10^6 and 5 x 10^6, preferably between about 0.5 x 10^6 and 2 x 10^6. A number of suitable viral and nonviral-based gene transfer vectors have been described for use herein.

After transduction, transduced cells are selected away from non-transduced cells using known techniques. For example, if the gene transfer vector used in the transduction includes a selectable marker which confers resistance to a cytotoxic agent, the cells can be contacted with the appropriate cytotoxic agent, whereby non-transduced cells can be negatively selected away from the transduced cells. If the selectable marker is a cell surface marker, the cells can be contacted with a binding agent specific for the particular cell surface marker, whereby the transduced cells can be positively selected away from the population. The selection step can also entail fluorescence-activated cell sorting (FACS) techniques, such as where FACS is used to select cells from the population containing a particular surface marker, or the selection step can entail the use of magnetically responsive particles as retrievable supports for target cell capture and/or background removal.

More particularly, positive selection of the transduced cells can be performed using a FACS cell sorter (e.g. a FACSVantage™ Cell Sorter, Becton Dickinson Immunocytometry Systems, San Jose, CA) to sort and collect transduced cells expressing a selectable cell surface marker. Following transduction, the cells are stained with fluorescent-labeled antibody molecules directed against the particular cell surface marker. The amount of bound antibody on each cell can be measured by passing droplets containing the cells through the cell sorter. By imparting an electromagnetic charge to droplets containing the stained cells, the transduced cells can be separated from other cells. The positively selected cells are then harvested in sterile collection vessels. These cell sorting procedures are described in detail, for example, in the FACSVantage™ Training Manual, with particular reference to sections 3-11 to 3-28 and 10-1 to 10-17.

Positive selection of the transduced cells can also be performed using magnetic separation of cells based on expression or a particular cell surface marker. In such separation techniques, cells to be positively selected are first contacted with specific binding agent (e.g., an antibody or reagent the interacts specifically with the cell surface marker). The cells are then contacted with retrievable particles (e.g., magnetically responsive particles) which are coupled with a reagent that binds the specific binding agent (that has bound to the positive cells). The cell-binding agent-particle complex can then be physically separated from non-labeled cells, for example using a magnetic field. When using magnetically responsive particles, the labeled cells can be retained in a container using a magnetic filed while the

negative cells are removed. These and similar separation procedures are known to those of ordinary skill in the art.

Expression of the vector in the selected transduced cells can be assessed by a number of assays known to those skilled in the art. For example, Western blot or Northern analysis can be employed depending on the nature of the inserted nucleotide sequence of interest. Once expression has been established and the transformed T cells have been tested for the presence of the selected synthetic expression cassette, they are ready for infusion into a patient via the peripheral blood stream.

The invention includes a kit for genetic modification of an *ex vivo* population of primary mammalian cells. The kit typically contains a gene transfer vector coding for at least one selectable marker and at least one synthetic expression cassette contained in one or more containers, ancillary reagents or hardware, and instructions for use of the kit.

2.4.3 FURTHER DELIVERY REGIMES

Any of the polynucleotides (*e.g.*, expression cassettes) or polypeptides described herein (delivered by any of the methods described above) can also be used in combination with other DNA delivery systems and/or protein delivery systems. Non-limiting examples include co-administration of these molecules, for example, in prime-boost methods where one or more molecules are delivered in a "priming" step and, subsequently, one or more molecules are delivered in a "boosting" step. In certain embodiments, the delivery of one or more nucleic acid-containing compositions and is followed by delivery of one or more nucleic acid-containing compositions and/or one or more polypeptide-containing compositions (*e.g.*, polypeptides comprising HIV antigens). In other embodiments, multiple nucleic acid "primes" (of the same or different nucleic acid molecules) can be followed by multiple polypeptide "boosts" (of the same or different polypeptides). Other examples include multiple nucleic acid administrations and multiple polypeptide administrations.

In any method involving co-administration, the various compositions can be delivered in any order. Thus, in embodiments including delivery of multiple different compositions or molecules, the nucleic acids need not be all delivered before the polypeptides. For example, the priming step may include delivery of one or more polypeptides and the boosting comprises delivery of one or more nucleic acids and/or one more polypeptides. Multiple polypeptide administrations can be followed by multiple nucleic acid administrations or polypeptide and nucleic acid administrations can be performed in any order. In any of the embodiments described herein, the nucleic acid molecules can encode all, some or none of the polypeptides. Thus, one or more or the nucleic

acid molecules (*e.g.*, expression cassettes) described herein and/or one or more of the polypeptides described herein can be co-administered in any order and via any administration routes. Therefore, any combination of polynucleotides and/or polypeptides described herein can be used to generate elicit an immune reaction.

EXPERIMENTAL

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

Example 1

Generation of Synthetic Expression Cassettes

A. Modification of HIV-1 *Env, Gag, Pol* Nucleic Acid Coding Sequences

The Pol coding sequences were selected from Type C strain AF110975. The Gag coding sequences were selected from the Type C strains AF110965 and AF110967. The Env coding sequences were selected from Type C strains AF110968 and AF110975. These sequences were manipulated to maximize expression of their gene products.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences (Schneider R, et al., *J Virol.* 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements are inactivated by introducing multiple point mutations that do not alter the reading frame of the encoded proteins. Figures 5 and 6 (SEQ ID Nos: 3, 4, 20 and 21) show the location of some remaining INS in synthetic sequences derived from strains AF110965 and AF110967. The changes made to these sequences are boxed in the Figures. In Figures 5 and

6, the top line depicts a modified sequence of Gag polypeptides from the indicated strains. The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to further remove INS. Thus, when the changes indicated in the boxed regions are made, the resulting sequences correspond to the sequences depicted in Figures 1 and 2, respectively.

The synthetic coding sequences are assembled by methods known in the art, for example by companies such as the Midland Certified Reagent Company (Midland, Texas).

In one embodiment of the invention, sequences encoding Pol-polypeptides are included with the synthetic Gag or Env sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, modified Gag/Env expression cassette. Because synthetic HIV-1 Pol expresses the functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it may be helpful in some instances to inactivate RT and INT functions. Several deletions or mutations in the RT and INT coding regions can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York. ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* **9(1)**:58-68; Wakefield, J. K.,et al., (1992) *Journal Of Virology* **66(11)**:6806-6812; Esnouf, R.,et al., (1995) *Nature Structural Biology* **2(4)**:303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* **282(2)**:359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* **73** (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* **90(13)**:6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* **269(46)**:29279-29287; Goldgur, Y., et al., (1998) *Proceedings Of the National Academy Of Sciences Of the United States Of America* **95(16)**:9150-9154; Goette, M., et al., (1998) *Journal Of Biological Chemistry* **273(17)**:10139-10146; Gorton, J. L., et al., (1998) *Journal of Virology* **72(6)**:5046-5055; Engelman, A., et al., (1997) *Journal Of Virology* **71(5)**:3507-3514; Dyda, F., et al., *Science* **266(5193)**:1981-1986; Davies, J. F., et al., (1991) *Science* **252(5002)**:88-95; Bujacz, G., et al., (1996) *Febs Letters* **398(2-3)**:175-178; Beard, W. A., et al., (1996) *Journal Of Biological Chemistry* **271(21)**:12213-12220; Kohlstaedt, L. A., et al., (1992) *Science* **256(5065)**:1783-1790; Krug, M. S. and Berger, S. L. (1991) *Biochemistry* **30(44)**:10614-10623; Mazumder, A., et al., (1996) *Molecular Pharmacology* **49(4)**:621-628; Palaniappan, C., et al., (1997) *Journal Of Biological Chemistry* **272(17)**:11157-11164; Rodgers, D. W., et al., (1995) *Proceedings Of the National Academy Of Sciences Of the United States Of America* **92(4)**:1222-1226; Sheng, N. and Dennis, D. (1993) *Biochemistry* **32(18)**:4938-4942; Spence, R. A., et al., (1995) *Science* **267(5200)**:988-993.}

Furthermore selected B- and/or T-cell epitopes can be added to the Pol constructs (*e.g.*, 3' of the truncated INT or within the deletions of the RT- and INT-coding sequence) to replace and augment any epitopes deleted by the functional modifications of RT and INT. Alternately, selected B- and T-cell epitopes (including CTL epitopes) from RT and INT can be included in a minimal VLP formed by expression of the synthetic Gag or synthetic Pol cassette, described above. (For descriptions of known HIV B- and T-cell epitopes see, HIV Molecular Immunology Database CTL Search Interface; Los Alamos Sequence Compendia, 1987-1997;Internet address: http://hiv-web.lanl.gov/immunology/index.html.)

The resulting modified coding sequences are presented as a synthetic Env expression cassette; a synthetic Gag expression cassette; a synthetic Pol expression cassette. A common Gag region (Gag-common) extends from nucleotide position 844 to position 903 (SEQ ID NO:1), relative to AF110965 (or from approximately amino acid residues 282 to 301 of SEQ ID NO:17) and from nucleotide position 841 to position 900 (SEQ ID NO:2), relative to AF110967 (or from approximately amino acid residues 281 to 300 of SEQ ID NO:22). A common Env region (Env-common) extends from nucleotide position 1213 to position 1353 (SEQ ID NO:5) and amino acid positions 405 to 451 of SEQ ID NO:23, relative to AF110968 and from nucleotide position 1210 to position 1353 (SEQ ID NO:11) and amino acid positions 404-451 (SEQ ID NO:24), relative to AF110975.

The synthetic DNA fragments for Pol, Gag and Env are cloned into the following eucaryotic expression vectors: pCMVKm2, for transient expression assays and DNA immunization studies, the pCMVKm2 vector is derived from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) and comprises a kanamycin selectable marker, a ColE1 origin of replication, a CMV promoter enhancer and Intron A, followed by an insertion site for the synthetic sequences described below followed by a polyadenylation signal derived from bovine growth hormone -- the pCMVKm2 vector differs from the pCMV-link vector only in that a polylinker site is inserted into pCMVKm2 to generate pCMV-link; pESN2dhfr and pCMVPLEdhfr, for expression in Chinese Hamster Ovary (CHO) cells; and, pAcC13, a shuttle vector for use in the Baculovirus expression system (pAcC13, is derived from pAcC12 which is described by Munemitsu S., et al., *Mol Cell Biol.* **10(11)**:5977-5982, 1990).

Briefly, construction of pCMVPLEdhfr was as follows.

To construct a DHFR cassette, the EMCV IRES (internal ribosome entry site) leader was PCR-amplified from pCite-4a+ (Novagen, Inc., Milwaukee, WI) and inserted into pET-23d (Novagen, Inc., Milwaukee, WI) as an *Xba-Nco* fragment to give pET-EMCV. The *dhfr* gene was PCR-amplified from pESN2dhfr to give a product with a Gly-Gly-Gly-Ser spacer in place of the translation stop codon and inserted as an *Nco-Bam*H1 fragment to give pET-E-

DHFR. Next, the attenuated *neo* gene was PCR amplified from a pSV2Neo (Clontech, Palo Alto, CA) derivative and inserted into the unique *Bam*H1 site of pET-E-DHFR to give pET-E-DHFR/$Neo_{(m2)}$. Finally the bovine growth hormone terminator from pCDNA3 (Invitrogen, Inc., Carlsbad, CA) was inserted downstream of the *neo* gene to give pET-E-DHFR/$Neo_{(m2)}$BGHt. The EMCV-*dhfr/neo* selectable marker cassette fragment was prepared by cleavage of pET-E-DHFR/$Neo_{(m2)}$BGHt.

The CMV enhancer/promoter plus Intron A was transferred from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) as a *Hin*dIII-*Sal*1 fragment into pUC19 (New England Biolabs, Inc., Beverly, MA). The vector backbone of pUC19 was deleted from the NdeI to the SapI sites. The above described DHFR cassette was added to the construct such that the EMCV IRES followed the CMV promoter. The vector also contained an amp^r gene and an SV40 origin of replication.

B. Defining of the Major Homology Region (MHR) of HIV-1 p55Gag

The Major Homology Region (MHR) of HIV-1 p55 (Gag) is located in the p24-CA sequence of Gag. It is a conserved stretch of approximately 20 amino acids. The position in the wild type AF110965 Gag protein is from 282-301 (SEQ ID NO:25) and spans a region from 844-903 (SEQ ID NO:26) for the Gag DNA-sequence. The position in the synthetic Gag protein is also from 282-301 (SEQ ID NO:25) and spans a region from 844-903 (SEQ ID NO:1) for the synthetic Gag DNA-sequence. The position in the wild type and synthetic AF110967 Gag protein is from 281-300 (SEQ ID NO:27) and spans a region from 841-900 (SEQ ID NO:2) for the modified Gag DNA-sequence. Mutations or deletions in the MHR can severely impair particle production (Borsetti, A., et al., *J. Virol.* **72(11):**9313-9317, 1998; Mammano, F., et al., *J Virol* 68(8):4927-4936, 1994).

Percent identity to this sequence can be determined, for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5.

C. Defining of the Common Sequence Region of HIV-1 Env

The common sequence region (CSR) of HIV-1 Env is located in the C4 sequence of Env. It is a conserved stretch of approximately 47 amino acids. The position in the wild type and synthetic AF110968 Env protein is from approximately amino acid residue 405 to 451 (SEQ ID NO:28) and spans a region from 1213 to 1353 (SEQ ID NO:5) for the Env DNA-sequence. The position in the wild type and synthetic AF110975 Env protein is

from approximately amino acid residue 404 to 451 (SEQ ID NO:29) and spans a region from 1210 to 1353 (SEQ ID NO:11) for the Env DNA-sequence.

Percent identity to this sequence can be determined, for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5.

Various forms of the different embodiments of the invention, described herein, may be combined.

D. Exemplary HIV Sequences Derived from South African HIV Type C Strains

HIV coding sequences of novel Type C isolates were obtained. Polypeptide-coding sequences were manipulated to maximize expression of their gene products.

As described above, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Shown below in Table C are exemplary wild-type and synthetic sequences derived from a novel South African HIV Type C isolate, clone 8_5_TV1_C.ZA. Table D shows exemplary synthetic Env sequences derived from a novel South African HIV Type C isolate, clone 8_2_TV1_C.ZA. Table E shows wild-type and synthetic sequences derived from South African HIV Type C strain 12-5_1_TV2_C.ZA.

Table C

Name	SEQ ID	Description
C4_Env_TV1_C_ZA_opt short	46	synthetic sequence of short Env "common region"
C4_Env_TV1_C_ZA_opt	47	synthetic sequence of Env "common region"
C4_Env_TV1_C_ZA_wt	48	wild type 8_5_TV1_C.ZA Env sequence
Envgp160_TV1_C_ZAopt	49	synthetic Env gp160
Envgp160_TV1_C_ZAwt	50	wild type 8_5_TV1_C.ZA Env gp160 sequence
Gag_TV1_C_ZAopt	51	synthetic sequence of Gag
Gag_TV1_C_ZAwt	52	wild type 8_5_TV1_C.ZA Gag sequence
Gag_TV1_ZA_MHRopt	53	synthetic sequence of Gag major homology region
Gag_TV1_ZA_MHRwt	54	wild type 8_5_TV1_C.ZA Gag major homology region sequence
Nef_TV1_C_ZAopt	55	synthetic sequence of Nef
Nef_TV1_C_ZAwt	56	wild type 8_5_TV1_C.ZA Nef sequence
NefD125G_TV1_C_ZAopt	57	synthetic sequence of Nef, including mutation at position 125 resulting in non-functional gene product
p15RNaseH_TV1_C_ZAopt	58	synthetic sequence of RNAseH (p15 of Pol)
p15RNaseH_TV1_C_ZAwt	59	wild type 8_5_TV1_C.ZA RNAseH sequence
p31Int_TV1_C_ZAopt	60	synthetic sequence of Integrase (p31 of Pol)
p31Int_TV1_C_ZAwt	61	wild type 8_5_TV1_C.ZA Integrase sequence
Pol_TV1_C_ZAopt	62	synthetic sequence of Pol
Pol_TV1_C_ZAwt	63	wild type 8_5_TV1_C.ZA Pol sequence
Prot_TV1_C_ZAopt	64	synthetic sequence of Prot
Prot_TV1_C_ZAwt	65	wild type 8_5_TV1_C.ZA Prot sequence
Protina_TV1_C_ZAopt	66	synthetic sequence of Prot including mutation resulting in inactivation of protease

Protina_TV1_C_ZAwt	67	wild type 8_5_TV1_C.ZA Prot sequence, including mutation resulting in inactivation of protease.
ProtinaRTmut_TV1_C_ZAop	68	synthetic sequence of Prot and reverse transcriptase (RT), including mutation resulting in inactivation of protease and mutation resulting in inactivation of RT.
ProtinaRTmut_TV1_C_ZAwt	69	wild type 8_5_TV1_C.ZA Prot and RT, mutation resulting in inactivation of protease and mutation resulting in inactivation of RT.
ProtwtRTwt_TV1_C_ZAopt	70	synthetic sequences of Prot and RT
ProtwtRTwt_TV1_C_ZAwt	71	wild type 8_5_TV1_C.ZA Prot and RT
RevExon1_TV1_C_ZAopt	72	synthetic sequence of exon 1 of Rev
RevExon1_TV1_C_ZAwt	73	wild type 8_5_TV1_C.ZA of exon 1 of Rev
RevExon2_TV1_C_ZAopt-2	74	synthetic sequence of exon 2 of Rev
RevExon2_TV1_C_ZAwt	75	wild type 8_5_TV1_C.ZA of exon 2 of Rev
RT_TV1_C_ZAopt	76	synthetic sequence of RT
RT_TV1_C_ZAwt	77	wild type 8_5_TV1_C.ZA RT
RTmut_TV1_C_ZAopt	78	synthetic sequence of RT, including mutation resulting in inactivation of RT
RTmut_TV1_C_ZAwt	79	wild type 8_5_TV1_C.ZA RT, including mutation resulting in inactivation of RT
TatC22Exon1_TV1_C_ZAopt	80	synthetic sequence of exon 1 of Tat, including mutation resulting in non-functional Tat gene product
TatExon1_TV1_C_ZAopt	81	synthetic sequence of exon 1 of Tat
TatExon1_TV1_C_ZAwt	82	wild type 8_5_TV1_C.ZA exon 1 of Tat
TatExon2_TV1_C_ZAopt	83	synthetic sequence of exon 2 of Tat
TatExon2_TV1_C_ZAwt	84	wild type 8_5_TV1_C.ZA exon 2 of Tat
Vif_TV1_C_ZAopt	85	synthetic sequence of Vif
Vif_TV1_C_ZAwt	86	wild type 8_5_TV1_C.ZA Vif
Vpr_TV1_C_ZAopt	87	synthetic sequence of Vpr

Vpr_TV1_C_ZAwt	88	wild type 8_5_TV1_C.ZA Vpr
Vpu_TV1_C_ZAopt	89	synthetic sequence of Vpu
Vpu_TV1_C_ZAwt	90	wild type 8_5_TV1_C.ZA Vpu
revexon1_2 TV1 C ZAopt	91	synthetic sequence of exons 1 and 2 of Rev
RevExon1_2_TV1_C_ZAwt	92	wild type 8_5_TV1_C.ZA Rev (exons 1 and 2)
TatC22Exon1_2_TV1_C_ZAopt	93	synthetic sequence of exons 1 and 2 of Tat, including mutation in exon 1 resulting in non-functional Tat gene product
TatExon1_2_TV1_C_ZAopt	94	synthetic sequence of exons 1 and 2 of Tat
TatExon1_2_TV1_C_ZAwt	95	wild type 8_5_TV1_C.ZA Tat (exons 1 and 2)
NefD125G-Myr_TV1_C_ZAopt	96	synthetic sequence of Nef, including mutation eliminating myristoylation site.

Table D

Name	Seq Id	Description
gp120mod.TV1.delV2	119	synthetic sequence of Env gp120, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.delV2	120	synthetic sequence of Env gp140, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.mut7.delV2	121	synthetic sequence of Env gp140, including V2 deletion and mutation in cleavage site and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.delV1V2	122	synthetic sequence of Env gp160, including V1/V2 deletion and modified leader derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.delV2	123	synthetic sequence of Env gp160, including V2 deletion and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.mut7.delV2	124	synthetic sequence of Env gp160, including V2 deletion; a mutation in cleavage site; and modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp160mod.TV1.tpa1	125	synthetic sequence of Env gp160, TPA1 leader
gp160mod.TV1	126	synthetic sequence of Env gp160, including modified leader sequences derived from wild-type (8_2_TV1_C.ZA) sequences
gp160mod.TV1.wtLnative	127	synthetic sequence of Env gp160, including wild type 8_2_TV1_C.ZA (unmodified) leader
gp140.mod.TV1.tpa1	131	synthetic sequence of Env gp140, TPA1 leader
gp140mod.TV1	132	synthetic sequence of Env gp140, including modified leader sequences derived from wild-type 8_2_TV1_C.ZA sequences
gp140mod.TV1.wtLnative	133	synthetic sequence of Env gp120, including wild type 8_2_TV1_C.ZA (unmodified) leader sequence.

As noted above, Env-encoding constructs can be prepared using any of the full-length of gp160 constructs. For example, a gp140 form (SEQ ID NO:132) was made by truncating gp160 (SEQ ID NO:126) at nucleotide 2064; gp120 was made by truncating gp160 (SEQ ID

NO:126) at nucleotide 1551 (SEQ ID NO:126). Additional gp140 and gp120 forms can be made using the methods described herein. One or more stop codons are typically added (*e.g.*, nucleotides 2608 to 2610 of SEQ ID NO:126). Further, the wild-type leader sequence can be modified and/or replaced with other leader sequences (*e.g*, TPA1 leader sequences).

Thus, the polypeptide gp160 includes the coding sequences for gp120 and gp41. The polypeptide gp41 is comprised of several domains including an oligomerization domain (OD) and a transmembrane spanning domain (TM). In the native envelope, the oligomerization domain is required for the non-covalent association of three gp41 polypeptides to form a trimeric structure: through non-covalent interactions with the gp41 trimer (and itself), the gp120 polypeptides are also organized in a trimeric structure. A cleavage site (or cleavage sites) exists approximately between the polypeptide sequences for gp120 and the polypeptide sequences corresponding to gp41. This cleavage site(s) can be mutated to prevent cleavage at the site. The resulting gp140 polypeptide corresponds to a truncated form of gp160 where the transmembrane spanning domain of gp41 has been deleted. This gp140 polypeptide can exist in both monomeric and oligomeric (*i.e.* trimeric) forms by virtue of the presence of the oligomerization domain in the gp41 moiety. In the situation where the cleavage site has been mutated to prevent cleavage and the transmembrane portion of gp41 has been deleted the resulting polypeptide product is designated "mutated" gp140 (e.g., gp140.mut). As will be apparent to those in the field, the cleavage site can be mutated in a variety of ways. In the exemplary constructs described herein (*e.g.*, SEQ ID NO:121 and SEQ ID NO:124), the mutation in the gp120/gp41 cleavage site changes the wild-type amino acid sequence KRRVVQREKR (SEQ ID NO:129) to ISSVVQSEKS (SEQ ID NO:130).

In yet other embodiments, hypervariable region(s) were deleted, N-glycosylation sites were removed and/or cleavage sites mutated. Exemplary constructs having variable region deletions (V1 and/or V2), V2 deletes were constructed by deleting nucleotides from approximately 499 to approximately 593 (relative to SEQ ID NO:128) and V1/V2 deletes were constructed by deleting nucleotides from approximately 375 to approximately 602 (relative to SEQ ID NO:128). The relative locations of V1 and/or V2 regions can also be readily determined by alignment to the regions shown in Table A. Table E shows wild-type and synthetic sequences derived from South African HIV Type C strain 12-5_1_TV2_C.ZA.

Table E

Name	SEQ ID	Description
Envgp160_TV2_C_ZAopt	97	synthetic sequence of Env gp160

Envgp160_TV2_C_ZAwt	98	wild type 12-5_1_TV2_C.ZA Env gp160.
Gag_TV2_C_ZAopt	99	synthetic sequence of Gag
Gag_TV2_C_ZAwt	100	wild type 12-5_1_TV2_C.ZA Gag
Nef_TV2_C_ZAopt	101	synthetic sequence of Nef
Nef_TV2_C_ZAwt	102	wild type 12-5_1_TV2_C.ZA Nef
Pol_TV2_C_ZAopt	103	synthetic sequence of Pol
Pol_TV2_C_ZAwt	104	wild type 12-5_1_TV2_C.ZA of Pol
RevExon1_TV2_C_ZAopt	105	synthetic sequence of exon 1 of Rev
RevExon1_TV2_C_ZAwt	106	wild type 12-5_1_TV2_C.ZA of exon 1 of Rev
RevExon2_TV2_C_ZAopt	107	synthetic sequence of exon 2 of Rev
RevExon2_TV2_C_ZAwt	108	wild type 12-5_1_TV2_C.ZA of exon 2 of Rev
TatExon1_TV2_C_ZAopt	109	synthetic sequence of exon 1 of Tat
TatExon1_TV2_C_ZAwt	110	wild type 12-5_1_TV2_C.ZA of exon 1 of Tat
TatExon2_TV2_C_ZAopt	111	synthetic sequence of exon 2 of Tat
TatExon2_TV2_C_ZAwt	112	wild type 12-5_1_TV2_C.ZA of exon 2 of Tat
Vif_TV2_C_ZAopt	113	synthetic sequence of Vif
Vif_TV2_C_ZAwt	114	wild type 12-5_1_TV2_C.ZA of Vif
Vpr_TV2_C_ZAopt	115	synthetic sequence of Vpr
Vpr_TV2_C_ZAwt	116	wild type 12-5_1_TV2_C.ZA of Vpr
Vpu_TV2_C_ZAopt	117	synthetic sequence of Vpu
Vpu_TV2_C_ZAwt	118	wild type 12-5_1_TV2_C.ZA of Vpu

It will be readily apparent that sequences derived from any HIV type C stain or clone can modified as described herein in order to achieve desirable modifications in that strain. Additionally, polyproteins can be constructed by fusing in-frame two or more polynucleotide sequences encoding polypeptide or peptide products. Further, polycistronic coding sequences may be produced by placing two or more polynucleotide sequences encoding polypeptide products adjacent each other, typically under the control of one promoter, wherein each polypeptide coding sequence may be modified to include sequences for internal ribosome binding sites.

The sequences of the present invention, for example, the modified (synthetic) polynucleotide sequences encoding HIV polypeptides, may be modified by deletions, point mutations, substitutions, frame-shifts, and/or further genetic modifications (for example, mutations leading to inactivation of an activity associated with a polypeptide, e.g., mutations that inactivate protease, tat, or reverse transcriptase activity). Such modifications are taught generally in the art and may be applied in the context of the teachings of the present invention. For example, sites corresponding to the "Regions of the HIV Genome" listed in Table A may be modified in the corresponding regions of the novel sequences disclosed herein in order to achieve desirable modifications. Further, the modified (synthetic) polynucleotide sequences of the present invention can be combined for use, e.g., in an composition for generating an immune response in a subject, in a variety of ways, including but not limited to the following ways: multiple individual expression cassettes each comprising one polynucleotide sequence of the present invention (e.g., a gag-expression cassette, an env expression cassette, and a rev expression cassette, or a pol-expression cassette, a vif expression cassette, and a vpr expression cassette, etc.); polyproteins produced by in-frame fusions of multiple polynucleotides of the present invention, and polycistronic polynucleotides produced using multiple polynulcleotides of the present invention.

Example 2

Expression Assays for the Synthetic Coding Sequences

A. Type C HIV Coding Sequences

The wild-type Subtype C HIV coding (for example from AF110965, AF110967, AF110968, AF110975, as well as novel South African strains 8_5_TV1_C.ZA, 8_2_TV1_C.ZA and 12-5_1_TV2_C.ZA) sequences are cloned into expression vectors having the same features as the vectors into which the synthetic sequences are cloned.

Expression efficiencies for various vectors carrying the wild-type and synthetic sequences are evaluated as follows. Cells from several mammalian cell lines (293, RD, COS-7, and CHO; all obtained from the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209) are transfected with 2 μg of DNA in transfection reagent LT1 (PanVera Corporation, 545 Science Dr., Madison, WI). The cells are incubated for 5 hours in reduced serum medium (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium is then replaced with normal medium as follows: 293 cells, IMDM, 10% fetal calf

serum, 2% glutamine (BioWhittaker, Walkersville, MD); RD and COS-7 cells, D-MEM, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD); and CHO cells, Ham's F-12, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The cells are incubated for either 48 or 60 hours. Cell lysates are collected as described below in Example 3. Supernatants are harvested and filtered through 0.45 μm syringe filters. Supernatants are evaluated using the using 96-well plates coated with a murine monoclonal antibody directed against HIV antigen, for example a Coulter p24-assay (Coulter Corporation, Hialeah, FL, US). The HIV-1 antigen binds to the coated wells. Biotinylated antibodies against HIV recognize the bound antigen. Conjugated strepavidin-horseradish peroxidase reacts with the biotin. Color develops from the reaction of peroxidase with TMB substrate. The reaction is terminated by addition of 4N H_2SO_4. The intensity of the color is directly proportional to the amount of HIV antigen in a sample.

Synthetic HIV Type C expression cassettes provides dramatic increases in production of their protein products, relative to the native (wild-type Subtype C) sequences, when expressed in a variety of cell lines.

B. Signal Peptide Leader Sequences

The ability of various leader sequences to drive expression was tested by transfecting cells with wild type or synthetic Env-encoding expression cassettes operably linked to different leader sequences and evaluating expression of Env polypeptide by ELISA or Western Blot. The amino acid and nucleotide sequence of various signal peptide leader sequences are shown in Table 4.

Table 4

Leader	Amino acid sequence	DNA sequence
WTnative (8_2_TV 1_C.ZA)	MRVMGTQKNCQQWWIWGI LGFWMLMIC	ATGAGAGTGATGGGGACACAGA AGAATTGTCAACAATGGTGGATA TGGGGCATCTTAGGCTTCTGGAT GCTAATGATTTGT
WTmod (8_2_TV 1_C.ZA)	MRVMGTQKNCQQWWIWGI LGFWMLMIC	ATGCGCGTGATGGGCACCCAGAA GAACTGCCAGCAGTGGTGGATCT GGGGCATCCTGGGCTTCTGGATG CTGATGATCTGC
Tpa1	MDAMKRGLCCVLLLCGAVFVSPS AS	ATGGATGCAATGAAGAGAGGGC TCTGCTGTGTGCTGCTGCTGTGTG

		GAGCAGTCTTCGTTTCGCCCAGC GCCAGC
Tpa2	MDAMKRGLCCVLLLCGAVFVSPS	ATGGATGCAATGAAGAGAGGGC TCTGCTGTGTGCTGCTGCTGTGTG GAGCAGTCTTCGTTTCGCCCAGC

293 cells were transiently transfected using standard methods with native and sequence-modified constructs encoding the gp120 and gp140 forms of the 8_2_TV1_C.ZA (TV1c8.2) envelope. Env protein was measure in cell lysates and supernatants using an in-house Env capture ELISA. Results are shown in Table 5 below and indicate that the wild-type signal peptide leader sequence of the TV1c8.2 can be used to efficiently express the encoded envelope protein to levels that are better or comparable to those observed using the heterologous tpa leader sequences. Furthermore, the TV1c8.2 leader works in its native or sequence-modified forms and can be used with native or sequence-modified env genes. All constructs were tested after cloning of the gene cassettes into the EcoR1 and Xho1 sites of the pCMVlink expression vector.

Table 5

TV1c8.2 construct	Supernatant (ng)	Lysate (ng)	Total (ng)
gp140nat.wtL	532	149	681
gp140nat.tpa1	250	20	270
gp140nat.tpa2	192	34	226
gp120mod.wtLmod	6186	4576	10762
gp120mod.tpa1	6932	3808	10740
gp120mod.wtLnat	6680	4174	10854
gp140mod.wtLmod	1844	8507	10351
gp140mod.tpa1	1854	2925	4779
gp140mod.wtLnat	1532	3015	4547

The sequence-modified TV1c8.2 envelope variant gene cassettes were subcloned into a Chiron pCMV expression vector for the derivation of stable mammalian cell lines. Stable CHO cell lines expressing the TV1c8.2 envelope proteins were derived using standard methods of transfection, methotrexate amplification, and screening. These cell lines were found to secrete levels of envelope protein that were comparable to those observed for proteins expressed using the tpa leader sequences. Representative results are shown in Table 6 for two cell line clone expressing the TV1c8.2 gp120; they are compared to two reference clones expressing SF162 subtype B gp120 derived in a similar fashion but using the tpa leader. Protein concentrations were determined following densitometry of scanned gels of semi-purified proteins. Standard curves were generated using a highly purified and well-characterized preparation of SF2 gp120 protein and the concentrations of the test proteins were determined.

Table 6

CHO cell line	Clone #	Expression (ng/ml)
gp120 SF162	Clone 65	921
	Clone 71	972
gp120TV1.C8.2	Clone 159	1977
	Clone 210	1920

The results were also confirmed by Western Blot Analysis, essentially as described in Example 3.

Example 3

Western Blot Analysis of Expression

A. HIV Type C Coding Sequences

Human 293 cells are transfected as described in Example 2 with pCMV-based vectors containing native or synthetic HIV Type C expression cassettes. Cells are cultivated for 60 hours post-transfection. Supernatants are prepared as described. Cell lysates are prepared as follows. The cells are washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO) in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) are loaded with 20 µl of supernatant or 12.5 µl of cell lysate. A protein standard is also loaded (5 µl, broad size range standard; BioRad Laboratories, Hercules, CA). Electrophoresis is carried out and the proteins are transferred using a BioRad Transfer Chamber (BioRad Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer (Millipore), where the transfer is performed at 100 volts for 90 minutes. The membranes are exposed to HIV-1-positive human patient serum and immunostained using o-phenylenediamine dihydrochloride (OPD; Sigma).

Immunoblotting analysis shows that cells containing the synthetic expression cassette produce the expected protein at higher per-cell concentrations than cells containing the native expression cassette. The proteins are seen in both cell lysates and supernatants. The levels of production are significantly higher in cell supernatants for cells transfected with the synthetic expression cassettes of the present invention.

In addition, supernatants from the transfected 293 cells are fractionated on sucrose gradients. Aliquots of the supernatant are transferred to Polyclear™ ultra-centrifuge tubes (Beckman Instruments, Columbia, MD), under-laid with a solution of 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 28,000 rpm in a Beckman SW28 rotor. The resulting pellet is suspended in PBS and layered onto a 20-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor.

The gradient is then fractionated into approximately 10 x 1 ml aliquots (starting at the top, 20%-end, of the gradient). Samples are taken from fractions 1-9 and are electrophoresed on 8-16% SDS polyacrylamide gels. The supernatants from 293/synthetic cells give much stronger bands than supernatants from 293/native cells.

Example 4

In Vivo Immunogenicity of Synthetic HIV Type C Expression Cassettes

A. Immunization

To evaluate the possibly improved immunogenicity of the synthetic HIV Type C expression cassettes, a mouse study is performed. The plasmid DNA, pCMVKM2 carrying the synthetic Gag expression cassette, is diluted to the following final concentrations in a total injection volume of 100 µl: 20 µg, 2 µg, 0.2 µg, 0.02 and 0.002 µg. To overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample is brought up to 20 µg using the vector (pCMVKM2) alone. As a control, plasmid DNA of the native Gag expression cassette is handled in the same manner. Twelve groups of four to ten Balb/c mice (Charles River, Boston, MA) are intramuscularly immunized (50 µl per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 1.

Table 1

Group	Gag or Env Expression Cassette	Concentration of Gag or Env plasmid DNA (μg)	Immunized at time (weeks):
1	Synthetic	20	0[1], 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	0.002	0, 4
6	Synthetic	20	0
7	Synthetic	2	0
8	Synthetic	0.2	0
9	Synthetic	0.02	0
10	Synthetic	0.002	0
11	Native	20	0, 4
12	Native	2	0, 4
13	Native	0.2	0, 4
14	Native	0.02	0, 4
15	Native	0.002	0, 4
16	Native	20	0
17	Native	2	0
18	Native	0.2	0
19	Native	0.02	0
20	Native	0.002	0

1 = initial immunization at "week 0"

Groups 1-5 and 11-15 are bled at week 0 (before immunization), week 4, week 6, week 8, and week 12. Groups 6-20 and 16-20 are bled at week 0 (before immunization) and at week 4.

B. Humoral Immune Response

The humoral immune response is checked with an anti-HIV antibody ELISAs (enzyme-linked immunosorbent assays) of the mice sera 0 and 4 weeks post immunization (groups 5-12) and, in addition, 6 and 8 weeks post immunization, respectively, 2 and 4 weeks post second immunization (groups 1-4).

The antibody titers of the sera are determined by using the appropriate anti-HIV polypeptide (*e.g.*, anti-Pol, anti-Gag, anti-Env, anti-Vif, anti-Vpu, etc.) antibody ELISA. Briefly, sera from immunized mice are screened for antibodies directed against the HIV proteins (*e.g.*, p55 Gag protein, an Env protein, *e.g.*, gp160 or gp120 or a Pol protein, *e.g.*, p6, prot or RT, etc). ELISA microtiter plates are coated with 0.2 μg of HIV protein per well overnight and washed four times; subsequently, blocking is done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100 μl of diluted mouse serum is added. Sera are tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates are washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates are washed and 100 μl of 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) is added per well. The optical density of each well is measured after 15 minutes. The titers reported are the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.).

Synthetic expression cassettes will provide a clear improvement of immunogenicity relative to the native expression cassettes.

C. Cellular Immune Response

The frequency of specific cytotoxic T-lymphocytes (CTL) is evaluated by a standard chromium release assay of peptide pulsed mouse (Balb/c, CB6F1 and/or C3H) CD4 cells. HIV polypeptide (*e.g.*, Pol, Gag or Env) expressing vaccinia virus infected CD-8 cells are used as a positive control. Briefly, spleen cells (Effector cells, E) are obtained from the mice immunized as described above are cultured, restimulated, and assayed for CTL activity against Gag peptide-pulsed target cells as described (Doe, B., and Walker, C.M., *AIDS* **10**(7):793-794, 1996). Cytotoxic activity is measured in a standard ^{51}Cr release assay. Target (T) cells are cultured with effector (E) cells at various E:T ratios for 4 hours and the average cpm from duplicate wells are used to calculate percent specific ^{51}Cr release.

Cytotoxic T-cell (CTL) activity is measured in splenocytes recovered from the mice immunized with HIV Gag or Env DNA. Effector cells from the Gag or Env DNA-immunized animals exhibit specific lysis of HIV polypeptide-pulsed SV-BALB (MHC matched) targets cells, indicative of a CTL response. Target cells that are peptide-pulsed and derived from an MHC-unmatched mouse strain (MC57) are not lysed.

Thus, synthetic expression cassettes exhibit increased potency for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

Example 5

DNA-immunization of Non-Human Primates Using a Synthetic HIV Type C Expression Cassette

Non-human primates are immunized multiple times (*e.g.,* weeks 0, 4, 8 and 24) intradermally, mucosally or bilaterally, intramuscular, into the quadriceps using various doses (*e.g.,* 1-5 mg) and various combinations of synthetic HIV Type C plasmids. The animals are bled two weeks after each immunization and ELISA is performed with isolated plasma. The ELISA is performed essentially as described in Example 4 except the second antibody-conjugate is an anti-human IgG, g-chain specific, peroxidase conjugate (Sigma Chemical Co., St. Louis, MD 63178) used at a dilution of 1:500. Fifty µg/ml yeast extract is added to the dilutions of plasma samples and antibody conjugate to reduce non-specific background due to preexisting yeast antibodies in the non-human primates.

Further, lymphoproliferative responses to antigen can also be evaluated post-immunization, indicative of induction of T-helper cell functions.

Synthetic plasmid DNA are expected to be immunogenic in non-human primates.

Example 6

In vitro expression of recombinant Sindbis RNA and DNA containing the synthetic HIV Type C expression cassette

To evaluate the expression efficiency of the synthetic Pol, Env and Gag expression cassette in Alphavirus vectors, the selected synthetic expression cassette is subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors. Specifically, a cDNA vector construct for *in vitro* transcription of Sindbis virus RNA vector replicons (pRSIN-luc; Dubensky, et al., *J Virol.* 70:508-519, 1996) is modified to contain a *Pme*I site for plasmid linearization and a polylinker for insertion of heterologous genes. A polylinker is generated using two oligonucleotides that contain the sites *Xho*I, *Pml*I, *Apa*I, *Nar*I, *Xba*I, *and Not*I (XPANXNF, and XPANXNR).

The plasmid pRSIN-luc (Dubensky et al., *supra)* is digested with *Xho*I and *Not*I to remove the luciferase gene insert, blunt-ended using Klenow and dNTPs, and purified from

an agarose get using GeneCleanII (Biol0l, Vista, CA). The oligonucleotides are annealed to each other and ligated into the plasmid. The resulting construct is digested with *NotI* and *SacI* to remove the minimal Sindbis 3'-end sequence and A_{40} tract, and ligated with an approximately 0.4 kbp fragment from PKSSIN1-BV (WO 97/38087). This 0.4 kbp fragment is obtained by digestion of pKSSIN1-BV with *NotI* and *Sac*I, and purification after size fractionation from an agarose gel. The fragment contains the complete Sindbis virus 3'-end, an A_{40} tract and a *Pme*I site for linearization. This new vector construct is designated SINBVE.

The synthetic HIV coding sequences are obtained from the parental plasmid by digestion with *EcoR*I, blunt-ending with Klenow and dNTPs, purification with GeneCleanII, digestion with *Sal*I, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic HIV polypeptide-coding fragment is ligated into the SINBVE vector that is digested with *Xho*I and *Pmt*I. The resulting vector is purified using GeneCleanII and is designated SINBVGag. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra)* from SINBVGag and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line.

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) is digested with *Sal*I and *Xba*I, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Gag or Env gene is inserted into the pDCMVSIN-beta-gal by digestion of SINBVGag with *Sal*I and *Xho*I, purification using GeneCleanII of the Gag-containing fragment after agarose gel size fractionation, and ligation. The resulting construct is designated pDSIN-Gag, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells are transfected with recombinant Sindbis RNA and DNA, respectively. The supernatants and cell lysates are tested with the Coulter capture ELISA (Example 2).

BHK cells are transfected by electroporation with recombinant Sindbis RNA.

293 cells are transfected using LT-1 (Example 2) with recombinant Sindbis DNA. Synthetic Gag- and/or Env-containing plasmids are used as positive controls. Supernatants and lysates are collected 48h post transfection.

Type C HIV proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic expression cassettes.

Example 7

In Vivo Immunogenicity of recombinant Sindbis Replicon Vectors containing synthetic Pol, Gag and/or Env Expression Cassettes

A. Immunization

To evaluate the immunogenicity of recombinant synthetic HIV Type C expression cassettes in Sindbis replicons, a mouse study is performed. The Sindbis virus DNA vector carrying synthetic expression cassettes (Example 6), is diluted to the following final concentrations in a total injection volume of 100 µl: 20 µg, 2 µg, 0.2 µg, 0.02 and 0.002 µg. To overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample is brought up to 20 µg using the Sindbis replicon vector DNA alone. Twelve groups of four to ten Balb/c mice (Charles River, Boston, MA) are intramuscularly immunized (50 µl per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 2. Alternatively, Sindbis viral particles are prepared at the following doses: 10^3 pfu, 10^5 pfu and 10^7 pfu in 100 µl, as shown in Table 3. Sindbis HIV polypeptide particle preparations are administered to mice using intramuscular and subcutaneous routes (50 µl per site).

Table 2

Group	Gag or Env Expression Cassette	Concentration of Gag or Env DNA (μg)	Immunized at time (weeks):
1	Synthetic	20	0[1], 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	0.002	0, 4
6	Synthetic	20	0
7	Synthetic	2	0
8	Synthetic	0.2	0
9	Synthetic	0.02	0
10	Synthetic	0.002	0

1 = initial immunization at "week 0"

Table 3

Group	Gag or Env sequence	Concentration of viral particle (pfu)	Immunized at time (weeks):
1	Synthetic	10^3	0[1], 4
2	Synthetic	10^5	0, 4
3	Synthetic	10^7	0, 4
8	Synthetic	10^3	0
9	Synthetic	10^5	0
10	Synthetic	10^7	0

1 = initial immunization at "week 0"

Groups are bled and assessment of both humoral and cellular (*e.g.*, frequency of specific CTLs) is performed, essentially as described in Example 4.

Example 8

Identification and Sequencing of a Novel HIV Type C Variants

A full-length clone, called 8_5_TV1_C.ZA, encoding an HIV Type C was isolated and sequenced. Briefly, genomic DNA from HIV-1 subtype C infected South African patients was isolated from PBMC (peripheral blood mononuclear cells) by alkaline lysis and anion-exchange columns (Quiagen). To get the genome of full-length clones two halves were amplified, that could later be joined together in frame within the Pol region using an unique Sal 1 site in both fragments. For the amplification, 200-800 ng of genomic DNA were added to the buffer and enzyme mix of the Expand Long Template PCR System after the protocol of the manufacturer (Boehringer Mannheim). The primer were designed after alignments of known full length sequences. For the 5'half a primer mix of 2 forward primers containing either thymidine (S1FCSacTA 5'-GTTTCTTGAGCTCTGGAAGGGTTAATTTAC TCCAAGAA-3', SEQ ID NO:38) or cytosine on position 20 (S1FTSacTA 5'-GTTTCTTGAGCTCTGGAAGGGTTAATTTACTCTAAGAA, SEQ ID NO:39) plus Sal 1 site, were used. The reverse primer were also a mix of two primers with either thymidine or cytosine on position 13 (S145RTSalTA 5'-GTTTCTTGTCGACTTGTCCATGTATGGCTTCCCC T-3', SEQ ID NO:40 and S145RCSalTA 5'-GTTTCTTGTCGACTTGTCCATGCATGGCTTCCCT-3' SEQ ID NO:41) and contained a Sal 1 site. The forward primer for the 3'half was also a mixture of two primers (S245FASalTA 5'-GTTTCTTGTCGACTGTAGTCCAGGaATATGGCAAT TAG-3' SEQ ID NO:42 and S245FGSalTA 5'-GTTTCTTGTCGACTGTAGTCCAGGgATATG GCAA TTAG-3' SEQ ID NO:43) with Sal 1 site and adenine or guanine on position 12. The reverse primer had a Not 1 site (S2_FullNotTA 5'-GTTTCTTGCGGCCGCTGCTAGA GATTTTCCACACTACCA-3' SEQ ID NO:44). After amplification the PCR products were purified using a 1% agarose gel and cloned into the pCR-XL-TOPO vector via TA cloning (Invitrogen). Colonies were checked by restriction analysis and sequence verified. For the full length sequence the sequences of the 5'- and 3'half were combined. The sequence is shown in SEQ ID NO:33. Furthermore, important domains are shown in Table A.

Another clone, designated 12-5_1_TV2_C.ZA was also sequenced and is shown in SEQ ID NO:45. The domains can be readily determined in view of the teachings of the specification, for example by aligning the sequence to those shown in Table A to find the corresponding regions in clone 12-5_1_TV2_C.ZA.

As described above (Example 1, Table C), synthetic expression cassettes were generated using one or more polynucleotide sequences obtained from 8_5_TV1_C.ZA or 12-5_1_TV2_C.ZA.

The polynucleotides described herein have all been deposited at Chiron Corporation, Emeryville, CA.

Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

Claims

1. An expression cassette comprising

a polynucleotide sequence encoding a polypeptide including an HIV *Pol* polypeptide, wherein the polynucleotide sequence encoding said *Pol* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented of Figure 8 (SEQ ID NO:30); Figure 9 (SEQ ID NO:31) or Figure 10 (SEQ ID NO:32).

2. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:46, (ii) X equals Y, and (iii) Y is at least 97.

3. The expression cassette of claim 2, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:47, (ii) X equals Y, and (iii) Y is at least 144.

4. The expression cassette of claim 3, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:49 or SEQ ID NO:97, (ii) X equals Y, and (iii) Y is at least 300.

5. The expression cassette of claim 4, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:49, (ii) X equals Y, and (iii) Y is 2610.

6. The expression cassette of claim 4, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:97, (ii) X equals Y, and (iii) Y is 2565.

7. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:51 (ii) X equals Y, and (iii) Y is 1494.

8. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:99, (ii) X equals Y, and (iii) Y is 1491.

9. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:55; SEQ ID NO:57; SEQ ID NO:101; SEQ ID NO:96; SEQ ID NO:134 or SEQ ID NO:135, (ii) X equals Y, and (iii) Y is at least 60.

10. The expression cassette of claim 9, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:55; SEQ ID NO:57; SEQ ID NO:101; SEQ ID NO:96; SEQ ID NO:134 or SEQ ID NO:135, (ii) X equals Y, and (iii) Y is 624.

11. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:58; (ii) X equals Y, and (iii) Y is 354.

12. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:60; (ii) X equals Y, and (iii) Y is 876.

13. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:62; (ii) X equals Y, and (iii) Y is 3015.

14. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:103; (ii) X equals Y, and (iii) Y is 3009.

15. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:64 or SEQ ID NO:66; (ii) X equals Y, and (iii) Y is 297.

16. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:68, (ii) X equals Y, and (iii) Y is 1965.

17. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:70; (ii) X equals Y, and (iii) Y is 1977.

18. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:72 or SEQ ID NO:105, (ii) X equals Y, and (iii) Y is at least 30.

19. The expression cassette of claim 18, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:72 or SEQ ID NO:105; (ii) X equals Y, and (iii) Y is 75.

20. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:74 or SEQ ID NO:107, (ii) X equals Y, and (iii) Y is at least 30.

21. The expression cassette of claim 20, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:74 or SEQ ID NO:107; (ii) X equals Y, and (iii) Y is 246.

22. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:76; (ii) X equals Y, and (iii) Y is 1680.

23. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:78; (ii) X equals Y, and (iii) Y is 1668.

24. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:80, SEQ ID NO:81 or SEQ ID NO:109; (ii) X equals Y, and (iii) Y is 216.

25. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:83; (ii) X equals Y, and (iii) Y is 93.

26. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:111; (ii) X equals Y, and (iii) Y is 90.

27. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:85, or SEQ ID NO:113; (ii) X equals Y, and (iii) Y is 579.

28. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:87; (ii) X equals Y, and (iii) Y is 288.

29. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:115; (ii) X equals Y, and (iii) Y is 287.

30. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:89 or SEQ ID NO:117; (ii) X equals Y, and (iii) Y is at least 30.

31. The expression cassette of claim 30 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:89; (ii) X equals Y, and (iii) Y is 267.

32. The expression cassette of claim 30 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:117; (ii) X equals Y, and (iii) Y is 261.

33. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:91; (ii) X equals Y, and (iii) Y is at least 30.

34. The expression cassette of claim 33 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:91; (ii) X equals Y, and (iii) Y is 321.

35. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:93 or SEQ ID NO:94; (ii) X equals Y, and (iii) Y is 309.

36. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:96; (ii) X equals Y, and (iii) Y is at least 60.

37. The expression cassette of claim 36 comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:96; (ii) X equals Y, and (iii) Y is 624.

38. An expression cassette comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:119, SEQ ID NO:120; SEQ ID NO:121; SEQ ID NO:122; SEQ ID NO:123; SEQ ID NO:124; SEQ ID NO:125; SEQ ID NO:126; SEQ ID NO:127; SEQ ID NO:131; SEQ ID NO:132 or SEQ ID NO:133, (ii) X equals Y, and (iii) Y is at least 60.

39. The expression cassette of claim 38, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:119, SEQ ID NO:120; SEQ ID NO:121; SEQ ID NO:122; SEQ ID NO:123; SEQ ID NO:124; SEQ ID NO:125; SEQ ID NO:126; SEQ ID NO:127; SEQ ID NO:131; SEQ ID NO:132 or SEQ ID NO:133,(ii) X equals Y, and (iii) Y is at least 300.

40. The expression cassette of claim 39,comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:123 or SEQ ID NO:124, (ii) X equals Y, and (iii) Y is 2433.

41. The expression cassette of claim 39, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:122, (ii) X equals Y, and (iii) Y is 2301.

42. The expression cassette of claim 39, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:125; (ii) X equals Y, and (iii) Y is 2517.

43. The expression cassette of claim 39, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:126 or SEQ ID NO:127, (ii) X equals Y, and (iii) Y is 2520.

44. The expression cassette of claim 39, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:119, (ii) X equals Y, and (iii) Y is 1377.

45. The expression cassette of claim 39, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:120 or SEQ ID NO:121, (ii) X equals Y, and (iii) Y is 1839.

46. The expression cassette of claim 39, comprising

a polynucleotide comprising X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least 90% percent identity to Y contiguous nucleotides of SEQ ID NO:132 or SEQ ID NO:133, (ii) X equals Y, and (iii) Y is 1890.

47. A polynucleotide comprising the sequence depicted in SEQ ID NO:33 or fragments derived therefrom.

48. The polynucleotide of claim 47, wherein said fragments comprise coding sequence for the gene products selected from the group consisting of Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env and Nef.

49. The polynucleotide of claim 48, wherein the fragment comprises a Gag gene product.

50. The polynucleotide of claim 48, wherein the fragment comprises an Env gene product.

51. The polynucleotide of claim 50, wherein the Env gene product is gp160, gp140 or gp120.

52. A polynucleotide comprising the sequence depicted in SEQ ID NO:45 or fragments derived therefrom.

53. The polynucleotide of claim 52, wherein said fragments comprise coding sequence for the gene products selected from the group consisting of Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env and Nef.

54. The polynucleotide of claim 53, wherein the fragment comprises a Gag gene product.

55. The polynucleotide of claim 53, wherein the fragment comprises an Env gene product.

56. The polynucleotide of claim 55, wherein the Env gene product is gp160, gp140 or gp120.

57. A polynucleotide comprising the sequence depicted in SEQ ID NO:128 or fragments derived therefrom.

58. The polynucleotide of claim 57, wherein the fragments comprise coding sequence for Env gene products gp160, gp140 or gp120.

59. The expression cassette of any of claims 1 to 46, further comprising one or more nucleic acids encoding one or more viral polypeptides or antigens.

60. The expression cassette of claim 59, wherein the viral polypeptide or antigen is selected from the group consisting of Gag, Env, vif, vpr, tat, rev, vpu, nef and combinations thereof.

61. The expression cassette of any of claims 1 to 46, further comprising one or more nucleic acids encoding one or more cytokines.

62. A recombinant expression system for use in a selected host cell, comprising, an expression cassette of any of claims 1 to 46, and wherein said polynucleotide sequence further comprises control elements capable of driving expression in the selected host cell.

63. The recombinant expression system of claim 62, wherein said control elements are selected from the group consisting of a transcription promoter, a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

64. The recombinant expression system of claim 62 wherein said transcription promoter is selected from the group consisting of CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

65. A cell comprising an expression cassette of any of claims 1 to 46, and wherein said polynucleotide sequence further comprises control elements compatible with expression in the selected cell.

66. The cell of claim 65, wherein the cell is selected from the group consisting of a mammalian cell, an insect cell, a bacterial cell, a yeast cell, a plant, an antigen presenting cell, a primary cell, an immortalized cell, and a tumor derived cell.

67. The cell of claim 66, wherein the cell is selected from the group consisting of BHK, VERO, HT1080, 293, RD, COS-7, and CHO cells.

68. The cell of claim 67, wherein said cell is a CHO cell.

69. The cell of claim 66, wherein the cell is either *Trichoplusia ni* (Tn5) or Sf9 insect cells.

70. The cell of claim 66, wherein the antigen presenting cell is a lymphoid cell selected from the group consisting of macrophage, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof.

71. A composition for generating an immunological response, comprising an expression cassette of any of claims 1 to 46.

72. The composition of claim 71, further comprising one or more *Pol* polypeptides.

73. The composition of claim 72, further comprising an adjuvant.

74. A composition for generating an immunological response, comprising an expression cassette of claim 52.

75. The composition of claim 74, further comprising a *Pol* polypeptide.

76. The composition of claim 74, further comprising one or more polypeptides encoded by the nucleic acid molecules of claim 60.

77. The composition of claim 76, further comprising an adjuvant.

78. A method of immunization of a subject, comprising,

introducing a composition of claim 71 into said subject under conditions that are compatible with expression of said expression cassette in said subject.

79. The method of claim 78, wherein said expression cassette is introduced using a gene delivery vector.

80. The method of claim 79, wherein the gene delivery vector is a non-viral vector.

81. The method of claim 79, wherein said gene delivery vector is a viral vector.

82. The method of claim 79, wherein said gene delivery vector is selected from the group consisting of an adenoviral vector, a vaccinia viral vector, an AAV vector, a retroviral vector, a lentiviral vector and an alphaviral vector.

83. The method of claim 82, wherein said gene delivery vector is a Sindbis-virus derived vector.

84. The method of claim 82, wherein said gene delivery vector is a cDNA vector.

85. The method of claim 82, wherein said gene delivery vector is a eukaryotic layered viral initiation system (ELVIS).

86. The method of claim 79, wherein said composition delivered using a particulate carrier.

87. The method of claim 79, wherein said composition is coated on a gold or tungsten particle and said coated particle is delivered to said subject using a gene gun.

88. The method of claim 79, wherein said composition is encapsulated in a liposome preparation.

89. The method of claim 79, wherein said subject is a mammal.

90. The method of claim 89, wherein said mammal is a human.

91. A method of generating an immune response in a subject, comprising:
providing an expression cassette of any of claims 1 to 46,
expressing said polypeptide in a suitable host cell,
isolating said polypeptide, and
administering said polypeptide to the subject in an amount sufficient to elicit an immune response.

92. A method of generating an immune response in a subject, comprising introducing into cells of said subject an expression cassette of any one of claims 1 to 46, under conditions that permit the expression of said polynucleotide and production of said polypeptide, thereby eliciting an immunological response to said polypeptide.

93. The method of claim 92, where the method further comprises co-administration of an HIV polypeptide.

94. The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out before introducing said expression cassette.

95. The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out concurrently with introducing said expression cassette.

96. The method of claim 93, wherein co-administration of the polypeptide to the subject is carried out after introducing said expression cassette.

97. The expression cassette of claim 59, wherein the viral polypeptide or antigen is selected from the group consisting of polypeptides derived from hepatitis B, hepatitis C and combinations thereof.

Gag_AF110965_BW_mod

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGGCAAGCTGGACGCCTGGGAGCGCATCCGCC
TGCGCCCCGGCGGCAAGAAGTGCTACATGATGAAGCACCTGGTGTGGGCCAGCCGCGAGCT
GGAGAAGTTCGCCCTGAACCCCGGCCTGCTGGAGACCAGCGAGGGCTGCAAGCAGATCATC
CGCCAGCTGCACCCCGCCCTGCAGACCGGCAGCGAGGAGCTGAAGAGCCTGTTCAACACCG
TGGCCACCCTGTACTGCGTGCACGAGAAGATCGAGGTCCGCGACACCAAGGAGGCCCTGGA
CAAGATCGAGGAGGAGCAGAACAAGTGCCAGCAGAAGATCCAGCAGGCCGAGGCCGCCGAC
AAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCACC
AGGCCATCAGCCCCCGCACCCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAG
CCCCGAGGTGATCCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAAC
ACGATGTTGAACACCGTGGGCGGCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCA
ACGAGGAGGCCGCCGAGTGGGACCGCGTGCACCCCGTGCACGCCGGCCCCATCGCCCCCGG
CCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCACCACCAGCACCCTGCAGGAGCAG
ATCGCCTGGATGACCAGCAACCCCCCCATCCCCGTGGGCGACATCTACAAGCGGTGGATCA
TCCTGGGCCTGAACAAGATCGTGCGGATGTACAGCCCCGTGAGCATCCTGGACATCAAGCA
GGGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAG
CAGAGCACCCAGGAGGTGAAGAACTGGATGACCGACACCCTGCTGGTGCAGAACGCCAACC
CCGACTGCAAGACCATCCTGCGCGCTCTCGGCCCCGGCGCCAGCCTGGAGGAGATGATGAC
CGCCTGCCAGGGCGTGGGCGGCCCCAGCCACAAGGCCCGCGTGCTGGCCGAGGCGATGAGC
CAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTCAAGGGCCCCCGGCGCATCGTCA
AGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAA
GGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC
AACTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCCGGCAACTTCCTGCAGAGCC
GCCCCGAGCCCACCGCCCCCCCCGCCGAGAGCTTCCGCTTCGAGGAGACCACCCCCGGCCA
GAAGCAGGAGAGCAAGGACCGCGAGACCCTGACCAGCCTGAAGAGCCTGTTCGGCAACGAC
CCCCTGAGCCAGTAA

Figure 1

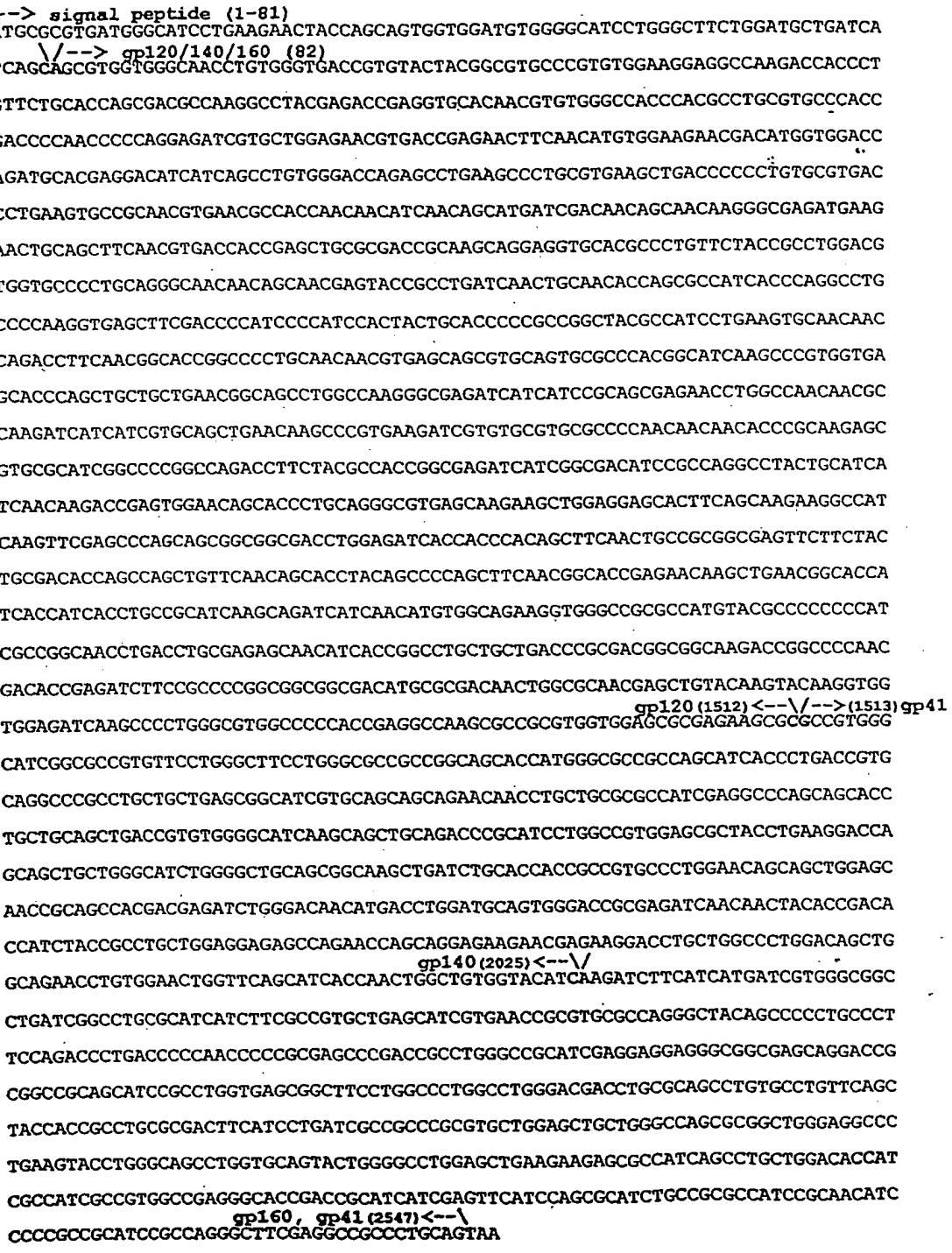
Gag_AF110967_BW_mod

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGAGAAGCTGGACAAGTGGGAGAAGATCCGCC
TGCGCCCCGGCGGCAAGAAGCACTACATGCTGAAGCACCTGGTGTGGGCCAGCCGCGAGCT
GGAGGGCTTCGCCCTGAACCCCGGCCTGCTGGAGACCGCCGAGGGCTGCAAGCAGATCATG
AAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGGAGCTGCGCAGCCTGTACAACACCG
TGGCCACCCTGTACTGCGTGCACGCCGGCATCGAGGTCCGCGACACCAAGGAGGCCCTGGA
CAAGATCGAGGAGGAGCAGAACAAGTCCCAGCAGAAGACCCAGCAGGCCAAGGAGGCCGAC
GGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCACCAGG
CCATCAGCCCCCGCACCCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCC
CGAGGTGATCCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAACACG
ATGTTGAACACCGTGGGCGGCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAACG
AGGAGGCCGCCGAGTGGGACCGCCTGCACCCCGTGCAGGCCGGCCCCGTGGCCCCCGGCCA
GATGCGCGACCCCCGCGGCAGCGACATCGCCGGCGCCACCAGCACCCTGCAGGAGCAGATC
GCCTGGATGACCAGCAACCCCCCCGTGCCCGTGGGCGACATCTACAAGCGGTGGATCATCC
TGGGCCTGAACAAGATCGTGCGGATGTACAGCCCCGTGAGCATCCTGGACATCCGCCAGGG
CCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAG
GCCACCCAGGACGTGAAGAACTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCG
ACTGCAAGACCATCCTGCGCGCTCTCGGCCCCGGCGCCACCCTGGAGGAGATGATGACCGC
CTGCCAGGGCGTGGGCGGCCCCGGCCACAAGGCCCGCGTGCTGGCCGAGGCGATGAGCCAG
GCCAACAGCGTGAACATCATGATGCAGAAGAGCAACTTCAAGGGCCCCCGGCGCAACGTCA
AGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCAAGAACTGCCGCGCCCCCCGCAAGAA
GGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC
AACTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCCGGCAACTTCCTGCAGAACC
GCAGCGAGCCCGCCGCCCCCACCGTGCCCACCGCCCCCCCCGCCGAGAGCTTCCGCTTCGA
GGAGACCACCCCCGCCCCCAAGCAGGAGCCCAAGGACCGCGAGCCCTACCGCGAGCCCCTG
ACCGCCCTGCGCAGCCTGTTCGGCAGCGGCCCCCTGAGCCAGTAA

Figure 2

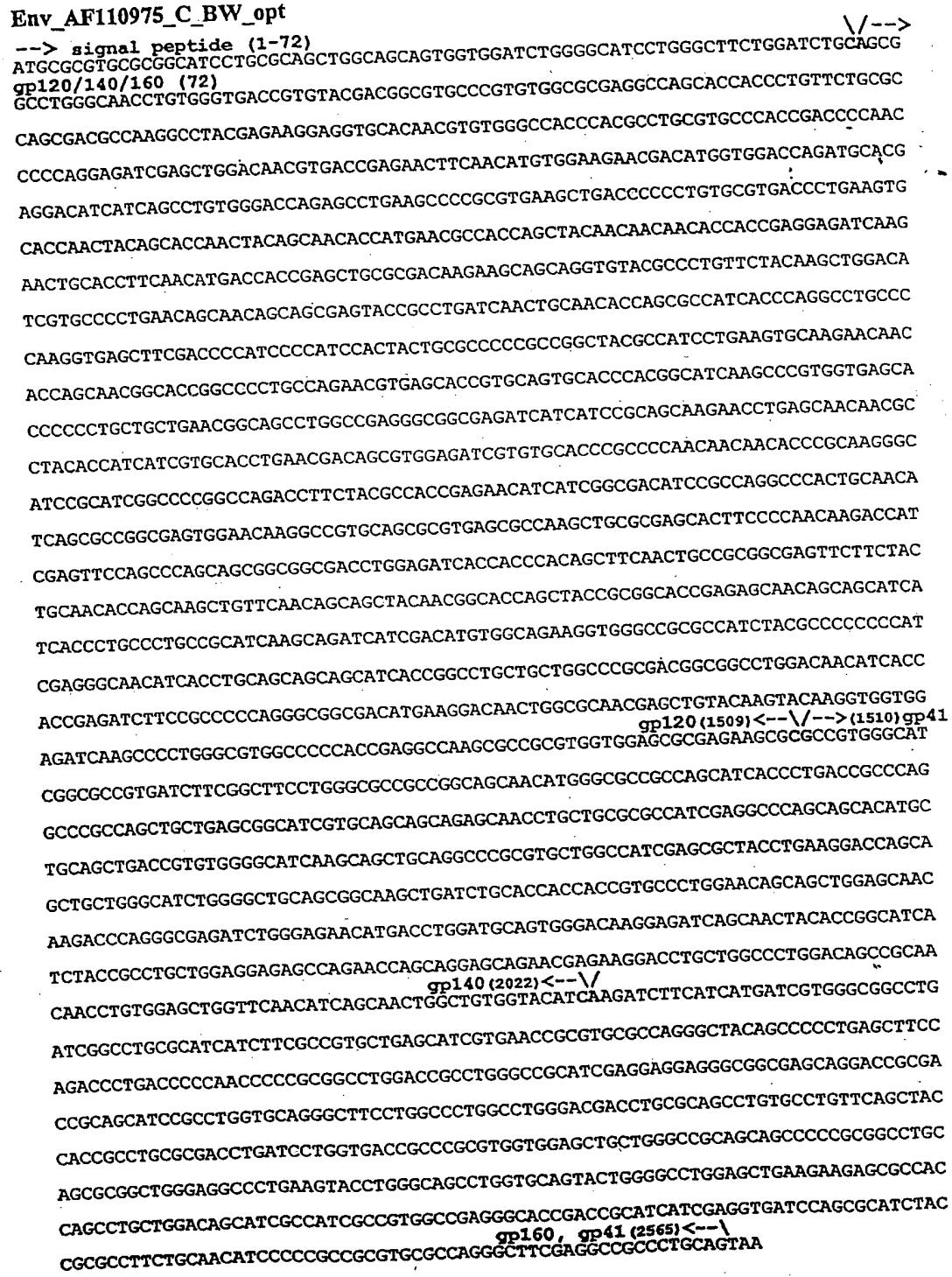
Fig. 3

Env_AF110968_C_BW_opt



ATGCGCGTGATGGGCATCCTGAAGAACTACCAGCAGTGGTGGATGTGGGGCATCCTGGGCTTCTGGATGCTGATCA
TCAGCAGCGTGGTGGGCAACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCAAGACCACCCT
GTTCTGCACCAGCGACGCCAAGGCCTACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCCACC
GACCCCAACCCCCAGGAGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGACC
AGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTGAC
CCTGAAGTGCCGCAACGTGAACGCCACCAACAACATCAACAGCATGATCGACAACAGCAACAAGGGCGAGATGAAG
AACTGCAGCTTCAACGTGACCACCGAGCTGCGCGACCGCAAGCAGGAGGTGCACGCCCTGTTCTACCGCCTGGACG
TGGTGCCCCTGCAGGGCAACAACAGCAACGAGTACCGCCTGATCAACTGCAACACCAGCGCCATCACCCAGGCCTG
CCCCAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCACCCCCGCCGGCTACGCCATCCTGAAGTGCAACAAC
CAGACCTTCAACGGCACCGGCCCCTGCAACAACGTGAGCAGCGTGCAGTGCGCCCACGGCATCAAGCCCGTGGTGA
GCACCCAGCTGCTGCTGAACGGCAGCCTGGCCAAGGGCGAGATCATCATCCGCAGCGAGAACCTGGCCAACAACGC
CAAGATCATCATCGTGCAGCTGAACAAGCCCGTGAAGATCGTGTGCGTGCGCCCCAACAACAACACCCGCAAGAGC
GTGCGCATCGGCCCCGGCCAGACCTTCTACGCCACCGGCGAGATCATCGGCGACATCCGCCAGGCCTACTGCATCA
TCAACAAGACCGAGTGGAACAGCACCCTGCAGGGCGTGAGCAAGAAGCTGGAGGAGCACTTCAGCAAGAAGGCCAT
CAAGTTCGAGCCCAGCAGCGGCGGCGACCTGGAGATCACCACCCACAGCTTCAACTGCCGCGGCGAGTTCTTCTAC
TGCGACACCAGCCAGCTGTTCAACAGCACCTACAGCCCCAGCTTCAACGGCACCGAGAACAAGCTGAACGGCACCA
TCACCATCACCTGCCGCATCAAGCAGATCATCAACATGTGGCAGAAGGTGGGCCGCGCCATGTACGCCCCCCCCAT
CGCCGGCAACCTGACCTGCGAGAGCAACATCACCGGCCTGCTGCTGACCCGCGACGGCGGCAAGACCGGCCCCAAC
GACACCGAGATCTTCCGCCCCGGCGGCGGCGACATGCGCGACAACTGGCGCAACGAGCTGTACAAGTACAAGGTGG
TGGAGATCAAGCCCCTGGGCGTGGCCCCCACCGAGGCCAAGCGCCGCGTGGTGGAGCGCGAGAAGCGCGCCGTGGG
CATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCACCATGGGCGCCGCCAGCATCACCCTGACCGTG
CAGGCCCGCCTGCTGCTGAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCCATCGAGGCCCAGCAGCACC
TGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCCGCATCCTGGCCGTGGAGCGCTACCTGAAGGACCA
GCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGGAACAGCAGCTGGAGC
AACCGCAGCCACGACGAGATCTGGGACAACATGACCTGGATGCAGTGGGACCGCGAGATCAACAACTACACCGACA
CCATCTACCGCCTGCTGGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAGAAGGACCTGCTGGCCCTGGACAGCTG
GCAGAACCTGTGGAACTGGTTCAGCATCACCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGC
CTGATCGGCCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCCTGCCCT
TCCAGACCCTGACCCCCAACCCCCGCGAGCCCGACCGCCTGGGCCGCATCGAGGAGGAGGGCGGCGAGCAGGACCG
CGGCCGCAGCATCCGCCTGGTGAGCGGCTTCCTGGCCCTGGCCTGGGACGACCTGCGCAGCCTGTGCCTGTTCAGC
TACCACCGCCTGCGCGACTTCATCCTGATCGCCGCCCGCGTGCTGGAGCTGCTGGGCCAGCGCGGCTGGGAGGCCC
TGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGCCTGGAGCTGAAGAAGAGCGCCATCAGCCTGCTGGACACCAT
CGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGTTCATCCAGCGCATCTGCCGCGCCATCCGCAACATC
CCCCGCCGCATCCGCCAGGGCTTCGAGGCCGCCCTGCAGTAA

Fig. 4


Env_AF110975_C_BW_opt
\/-->
--> signal peptide (1-72)
ATGCGCGTGCGCGGCATCCTGCGCAGCTGGCAGCAGTGGTGGATCTGGGGCATCCTGGGCTTCTGGATCTGCAGCG
gp120/140/160 (72)
GCCTGGGCAACCTGTGGGTGACCGTGTACGACGGCGTGCCCGTGTGGCGCGAGGCCAGCACCACCCTGTTCTGCGC
CAGCGACGCCAAGGCCTACGAGAAGGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCCACCGACCCCAAC
CCCCAGGAGATCGAGCTGGACAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGACCAGATGCACG
AGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCCGCGTGAAGCTGACCCCCCTGTGCGTGACCCTGAAGTG
CACCAACTACAGCACCAACTACAGCAACACCATGAACGCCACCAGCTACAACAACAACACCACCGAGGAGATCAAG
AACTGCACCTTCAACATGACCACCGAGCTGCGCGACAAGAAGCAGCAGGTGTACGCCCTGTTCTACAAGCTGGACA
TCGTGCCCCTGAACAGCAACAGCAGCGAGTACCGCCTGATCAACTGCAACACCAGCGCCATCACCCAGGCCTGCCC
CAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTACGCCATCCTGAAGTGCAAGAACAAC
ACCAGCAACGGCACCGGCCCCTGCCAGAACGTGAGCACCGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGAGCA
CCCCCCTGCTGCTGAACGGCAGCCTGGCCGAGGGCGGCGAGATCATCATCCGCAGCAAGAACCTGAGCAACAACGC
CTACACCATCATCGTGCACCTGAACGACAGCGTGGAGATCGTGTGCACCCGCCCCAACAACAACACCCGCAAGGGC
ATCCGCATCGGCCCCGGCCAGACCTTCTACGCCACCGAGAACATCATCGGCGACATCCGCCAGGCCCACTGCAACA
TCAGCGCCGGCGAGTGGAACAAGGCCGTGCAGCGCGTGAGCGCCAAGCTGCGCGAGCACTTCCCCAACAAGACCAT
CGAGTTCCAGCCCAGCAGCGGCGGCGACCTGGAGATCACCACCCACAGCTTCAACTGCCGCGGCGAGTTCTTCTAC
TGCAACACCAGCAAGCTGTTCAACAGCAGCTACAACGGCACCAGCTACCGCGGCACCGAGAGCAACAGCAGCATCA
TCACCCTGCCCTGCCGCATCAAGCAGATCATCGACATGTGGCAGAAGGTGGGCCGCGCCATCTACGCCCCCCCCAT
CGAGGGCAACATCACCTGCAGCAGCAGCATCACCGGCCTGCTGCTGGCCCGCGACGGCGGCCTGGACAACATCACC
ACCGAGATCTTCCGCCCCCAGGGCGGCGACATGAAGGACAACTGGCGCAACGAGCTGTACAAGTACAAGGTGGTGG
gp120 (1509)<--\/-->(1510) gp41
AGATCAAGCCCCTGGGCGTGGCCCCCACCGAGGCCAAGCGCCGCGTGGTGGAGCGCGAGAAGCGCGCCGTGGGCAT
CGGCGCCGTGATCTTCGGCTTCCTGGGCGCCGCCGGCAGCAACATGGGCGCCGCCAGCATCACCCTGACCGCCCAG
GCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAGAGCAACCTGCTGCGCGCCATCGAGGCCCAGCAGCACATGC
TGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGAAGGACCAGCA
GCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAAC
AAGACCCAGGGCGAGATCTGGGAGAACATGACCTGGATGCAGTGGGACAAGGAGATCAGCAACTACACCGGCATCA
TCTACCGCCTGCTGGAGGAGAGCCAGAACCAGCAGGAGCAGAACGAGAAGGACCTGCTGGCCCTGGACAGCCGCAA
gp140 (2022)<--\/
CAACCTGTGGAGCTGGTTCAACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGCCTG
ATCGGCCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCCTGAGCTTCC
AGACCCTGACCCCCAACCCCCGCGGCCTGGACCGCCTGGGCCGCATCGAGGAGGAGGGCGGCGAGCAGGACCGCGA
CCGCAGCATCCGCCTGGTGCAGGGCTTCCTGGCCCTGGCCTGGGACGACCTGCGCAGCCTGTGCCTGTTCAGCTAC
CACCGCCTGCGCGACCTGATCCTGGTGACCGCCCGCGTGGTGGAGCTGCTGGGCCGCAGCAGCCCCCGCGGCCTGC
AGCGCGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGCCTGGAGCTGAAGAAGAGCGCCAC
CAGCCTGCTGGACAGCATCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGATCCAGCGCATCTAC
gp160, gp41 (2565)<--\
CGCGCCTTCTGCAACATCCCCCGCCGCGTGCGCCAGGGCTTCGAGGCCGCCCTGCAGTAA

Gag_AF110965_BW_opt

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGGCAAGCTGGACGCCTGGGAGCGCATCCGCCTGCGCCCCGG

CGGCAAGAAGTGCTACATGATGAAGCACCTGGTGTGGGCCAGCCGCGAGCTGGAGAAGTTCGCCCTGAACC

CCGGCCTGCTGGAGACCAGCGAGGGCTGCAAGCAGATCATCCGCCAGCTGCACCCCGCCCTGCAGACCGGC

AGCGAGGAGCTGAAGAGCCTGTTCAACACCGTGGCCACCCTGTACTGCGTGCACGAGAAGATCGAGGTGCG
C

CGACACCAAGGAGGCCCTGGACAAGATCGAGGAGGAGCAGAACAAGTGCCAGCAGAAGATCCAGCAGGCCG

AGGCCGCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCAC

CAGGCCATCAGCCCCCGCACCCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCCCGAGGT

GATCCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAACACCATGCTGAACACCGTGG
G T

GCGGCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAACGAGGAGGCCGCCGAGTGGGACCGCGTG

CACCCCGTGCACGCCGGCCCCATCGCCCCCGGCCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCAC

CACCAGCACCCTGCAGGAGCAGATCGCCTGGATGACCAGCAACCCCCCCATCCCCGTGGGCGACATCTACA

AGCGCTGGATCATCCTGGGCCTGAACAAGATCGTGCGCATGTACAGCCCCGTGAGCATCCTGGACATCAAG
G G

CAGGGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAGAGCAC

CCAGGAGGTGAAGAACTGGATGACCGACACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCC

TGCGCGCCCTGGGCCCCGGCGCCAGCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCGGCCCCAGC
T C

CACAAGGCCCGCGTGCTGGCCGAGGCCATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTT
G

CAAGGGCCCCCCCCGCATCGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCG
G C

CCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAG

GCCAACTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCCGGCAACTTCCTGCAGAGCCGCCCCGA

GCCCACCGCCCCCCCCGCCGAGAGCTTCCGCTTCGAGGAGACCACCCCCGGCCAGAAGCAGGAGAGCAAGG

ACCGCGAGACCCTGACCAGCCTGAAGAGCCTGTTCGGCAACGACCCCCCTGAGCCAGTAA

Figure 5

Gag_AF110967_BW_opt

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGAGAAGCTGGACAAGTGGGAGAAGATCCGCCTGCGCCCCGG

CGGCAAGAAGCACTACATGCTGAAGCACCTGGTGTGGGCCAGCCGCGAGCTGGAGGGCTTCGCCCTGAACC

CCGGCCTGCTGGAGACCGCCGAGGGCTGCAAGCAGATCATGAAGCAGCTGCAGCCCGCCCTGCAGACCGGC

ACCGAGGAGCTGCGCAGCCTGTACAACACCGTGGCCACCCTGTACTGCGTGCACGCCGGCATCGAGGTGCG

CGACACCAAGGAGGCCCTGGACAAGATCGAGGAGGAGCAGAACAAGAGCAAGCAGCAGAAGACCCAGCAGGCCA

AGGAGGCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCACCAG

GCCATCAGCCCCCTGCACCCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCCCGAGGTGAT

CCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAACACCATGCTGAACACCGTGGGCG

GCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAACGAGGAGGCCGCCGAGTGGGACCGCCTGCAC

CCCGTGCAGGCCGGCCCCGTGGCCCCCGGCCAGATGCGCGACCCCCGCGGCAGCGACATCGCCGGCGCCAC

CAGCACCCTGCAGGAGCAGATCGCCTGGATGACCAGCAACCCCCCCGTGCCCGTGGGCGACATCTACAAGC

GCTGGATCATCCTGGGCCTGAACAAGATCGTGCGCATGTACAGCCCCGTGAGCATCCTGGACATCCGCCAG

GGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAGGCCACCCA

GGACGTGAAGAACTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCCTGC

GCGCCCTGGGCCCCGGCGCCACCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCGGCCCCGGCCAC

AAGGCCCGCGTGCTGGCCGAGGCCATGAGCCAGGCCAACAGCGTGAACATCATGATGCAGAAGAGCAACTT

CAAGGGCCCCCGCCGCAACGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCAAGAACTGCCGCG

CCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAG

GCCAACTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCCGGCAACTTCCTGCAGAACCGCAGCGA

GCCCGCCGCCCCCACCGTGCCCACCGCCCCCCCCGCCGAGAGCTTCCGCTTCGAGGAGACCACCCCCGCCC

CCAAGCAGGAGCCCAAGGACCGCGAGCCCTACCGCGAGCCCCTGACCGCCCTGCGCAGCCTGTTCGGCAGC

GGCCCCCTGAGCCAGTAA

Figure 6

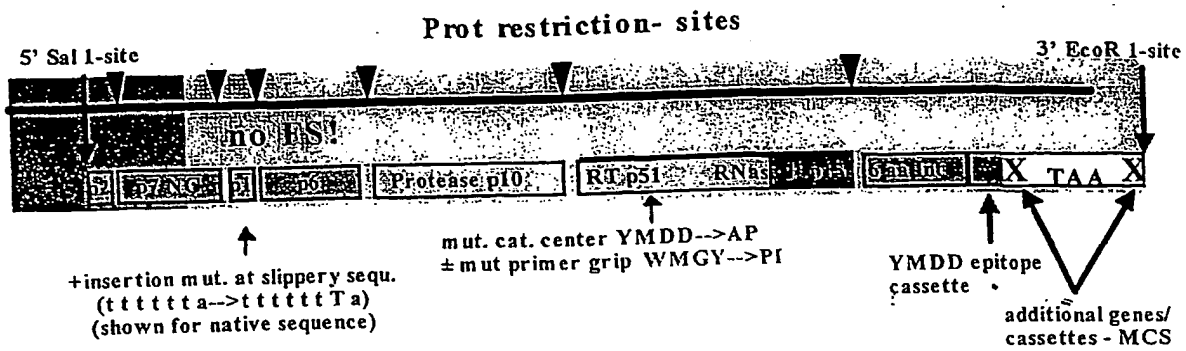

Prot restriction- sites
5' Sal 1-site
3' EcoR 1-site
no FS!
Protease p10
RT p51
RNas
X TAA X
+insertion mut. at slippery sequ.
(t t t t t t a-->t t t t t t T a)
(shown for native sequence)
mut. cat. center YMDD-->AP
± mut primer grip WMGY-->PI
YMDD epitope
cassette
additional genes/
cassettes - MCS

FIGURE 7

PR975(+) (SEQ ID NO:30)

GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGAT
GCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAA
GGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGT
GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTC
CGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAA
CCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCA
GCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGC
AGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC
ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC
CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCT
GATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT
GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCAT
CAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG
TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG
GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC
CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT
TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC
ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC
TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC
AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGC
TGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTC
CGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC
AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCT
GCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT
GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCC
CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACT
GGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG
GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTG
GCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAG
CAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA
TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACC
GCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGA
GAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGAC
CTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTT
CGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCAT
CGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCA
AGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACC
ACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAG
CGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCC
CGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG
AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAG
ATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT
GGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCT
AGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC

FIGURE 8

PR975YM (SEQ ID NO:31)

GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGAT
GCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAA
GGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGT
GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTC
CGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAA
CCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCA
GCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGC
AGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC
ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC
CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCT
GATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT
GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCAT
CAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG
TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG
GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC
CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT
TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC
ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC
TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC
AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGC
TGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTC
CGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGAC
CTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCG
CTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGAT
GGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA
AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCC
AGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCC
AAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGA
GAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGG
ACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC
CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCA
CACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCA
TCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGG
AGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGA
ACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG
CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCC
GGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA
CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG
TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACA
AGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAG
GTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGA
CAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCG
GCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGA
TCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC

FIGURE 9

PR975YMWM (SEQ ID NO:32)

GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGAT
GCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAA
GGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGT
GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTC
CGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAA
CCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCA
GCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGC
AGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGAC
ACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC
CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCT
GATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT
GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCAT
CAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG
TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAG
GAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC
CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT
TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC
ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCC
TACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCC
AGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGC
TGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTC
CGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGAC
CTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCG
CTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT
CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA
GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG
ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCC
CTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG
CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGT
GGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC
CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAAC
GACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGAT
CTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT
GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCC
CCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG
ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA
CGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA
AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAAC
ATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAG
CGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGT
ACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAG
CTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATC
GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGAT
TAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC

FIGURE 10

8_5_ZA (SEQ ID NO:33)

1 TGGAAGGGTT AATTTACTCC AAGAAAAGGC AAGAAATCCT TGATTTGTGG GTCTATCACA
61 CACAAGGCTT CTTCCCTGAT TGGCAAAACT ACACACCGGG GCCAGGGGTC AGATATCCAC
121 TGACCTTTGG ATGGTGCTAC AAGCTAGTGC CAGTTGACCC AGGGGAGGTG GAAGAGGCCA
181 ACGGAGGAGA AGACAACTGT TTGCTACACC CTATGAGCCA ACATGGAGCA GAGGATGAAG
241 ATAGAGAAGT ATTAAAGTGG AAGTTTGACA GCCTCCTAGC ACGCAGACAC ATGGCCCGCG
301 AGCTACATCC GGAGTATTAC AAAGACTGCT GACACAGAAG GGACTTTCCG CCTGGGACTT
361 TCCACTGGGG CGTTCCGGGA GGTGTGGTCT GGGCGGGACT TGGGAGTGGT CAACCCTCAG
421 ATGCTGCATA TAAGCAGCTG CTTTTCGCCT GTACTGGGTC TCTCTCGGTA GACCAGATCT
481 GAGCCTGGGA GCCCTCTGGC TATCTAGGGA ACCCACTGCT TAAGCCTCAA TAAAGCTTGC
541 CTTGAGTGCT TTAAGTAGTG TGTGCCCATC TGTTGTGTGA CTCTGGTAAC TAGAGATCCC
601 TCAGACCCTT TGTGGTAGTG TGGAAAATCT CTAGCAGTGG CGCCCGAACA GGGACCAGAA
661 AGTGAAAGTG AGACCAGAGG AGATCTCTCG ACGCAGGACT CGGCTTGCTG AAGTGCACAC
721 GGCAAGAGGC GAGAGGGGCG GCTGGTGAGT ACGCCAATTT TACTTGACTA GCGGAGGCTA
781 GAAGGAGAGA GATGGGTGCG AGAGCGTCAA TATTAAGCGG CGGAAAATTA GATAAATGGG
841 AAAGAATTAG GTTAAGGCCA GGGGGAAAGA AACATTATAT GTTAAAACAT CTAGTATGGG
901 CAAGCAGGGA GCTGGAAAGA TTTGCACTTA ACCCTGGCCT GTTAGAAACA TCAGAAGGCT
961 GTAAACAAAT AATAAAACAG CTACAACCAG CTCTTCAGAC AGGAACAGAG GAACTTAGAT
1021 CATTATTCAA CACAGTAGCA ACTCTCTATT GTGTACATAA AGGGATAGAG GTACGAGACA
1081 CCAAGGAAGC CTTAGACAAG ATAGAGGAAG AACAAAACAA ATGTCAGCAA AAAGCACAAC
1141 AGGCAAAAGC AGCTGACGAA AAGGTCAGTC AAAATTATCC TATAGTACAG AATGCCCAAG
1201 GGCAAATGGT ACACCAAGCT ATATCACCTA GAACATTGAA TGCATGGATA AAAGTAATAG
1261 AGGAAAAGGC TTTCAATCCA GAGGAAATAC CCATGTTTAC AGCATTATCA GAAGGAGCCA
1321 CCCCACAAGA TTTAAACACA ATGTTAAATA CAGTGGGGGG ACATCAAGCA GCCATGCAAA
1381 TGTTAAAAGA TACCATCAAT GAGGAGGCTG CAGAATGGGA TAGGACACAT CCAGTACATG
1441 CAGGGCCTGT TGCACCAGGC CAGATGAGAG AACCAAGGGG AAGTGACATA GCAGGAACTA
1501 CTAGTACCCT TCAGGAACAA ATAGCATGGA TGACAAGTAA TCCACCTATT CCAGTAGAAG
1561 ACATCTATAA AAGATGGATA ATTCTGGGGT TAAATAAAAT AGTAAGAATG TATAGCCCTG
1621 TTAGCATTTT GGACATAAAA CAAGGGCCAA AAGAACCCTT TAGAGACTAT GTAGACCGGT
1681 TCTTTAAAAC CTTAAGAGCT GAACAAGCTA CACAAGATGT AAAGAATTGG ATGACAGACA
1741 CCTTGTTGGT CCAAAATGCG AACCCAGATT GTAAGACCAT TTTAAGAGCA TTAGGACCAG
1801 GGGCCTCATT AGAAGAAATG ATGACAGCAT GTCAGGGAGT GGGAGGACCT AGCCATAAAG
1861 CAAGAGTGTT GGCTGAGGCA ATGAGCCAAG CAAACAGTAA CATACTAGTG CAGAGAAGCA
1921 ATTTTAAAGG CTCTAACAGA ATTATTAAAT GTTTCAACTG TGGCAAAGTA GGGCACATAG
1981 CCAGAAATTG CAGGGCCCCT AGGAAAAAGG GCTGTTGGAA ATGTGGACAG GAAGGACACC
2041 AAATGAAAGA CTGTACTGAG AGGCAGGCTA ATTTTTTAGG GAAAATTTGG CCTTCCCACA
2101 AGGGGAGGCC AGGGAATTTC CTCCAGAACA GACCAGAGCC AACAGCCCCA CCAGCAGAAC
2161 CAACAGCCCC ACCAGCAGAG AGCTTCAGGT TCGAGGAGAC AACCCCCGTG CCGAGGAAGG
2221 AGAAAGAGAG GGAACCTTTA ACTTCCCTCA AATCACTCTT TGGCAGCGAC CCCTTGTCTC
2281 AATAAAAGTA GAGGGCCAGA TAAAGGAGGC TCTCTTAGAC ACAGGAGCAG ATGATACAGT
2341 ATTAGAAGAA ATAGATTTGC CAGGGAAATG GAAACCAAAA ATGATAGGGG GAATTGGAGG
2401 TTTTATCAAA GTAAGACAGT ATGATCAAAT ACTTATAGAA ATTTGTGGAA AAAAGGCTAT
2461 AGGTACAGTA TTAGTAGGGC CTACACCAGT CAACATAATT GGAAGAAATC TGTTAACTCA
2521 GCTTGGATGC ACACTAAATT TTCCAATTAG TCCTATTGAA ACTGTACCAG TAAAATTAAA
2581 ACCAGGAATG GATGGCCCAA AGGTCAAACA ATGGCCATTG ACAGAAGAAA AAATAAAAGC
2641 ATTAACAGCA ATTTGTGAGG AAATGGAGAA GGAAGGAAAA ATTACAAAAA TTGGGCCTGA
2701 TAATCCATAT AACACTCCAG TATTTGCCAT AAAAAAGAAG GACAGTACTA AGTGGAGAAA
2761 ATTAGTAGAT TTCAGGGAAC TCAATAAAAG AACTCAAGAC TTTTGGGAAG TTCAATTAGG
2821 AATACCACAC CCAGCAGGAT TAAAAAAGAA AAAATCAGTG ACAGTGCTAG ATGTGGGGGA
2881 TGCATATTTT TCAGTTCCTT TAGATGAAAG CTTCAGGAAA TATACTGCAT TCACCATACC

FIGURE 11

```
2941 TAGTATAAAC AATGAAACAC CAGGGATTAG ATATCAATAT AATGTGCTGC CACAGGGATG
3001 GAAAGGATCA CCAGCAATAT TCCAGAGTAG CATGACAAAA ATCTTAGAGC CCTTCAGAGC
3061 AAAAAATCCA GACATAGTTA TCTATCAATA TATGGATGAC TTGTATGTAG GATCTGACTT
3121 AGAAATAGGG CAACATAGAG CAAAAATAGA AGAGTTAAGG GAACATTTAT TGAAATGGGG
3181 ATTTACAACA CCAGACAAGA AACATCAAAA AGAACCCCCA TTTCTTTGGA TGGGGTATGA
3241 ACTCCATCCT GACAAATGGA CAGTACAACC TATACTGCTG CCAGAAAAGG ATAGTTGGAC
3301 TGTCAATGAT ATACAGAAGT TAGTGGGAAA ATTAAACTGG GCAAGTCAGA TTTACCCAGG
3361 GATTAAAGTA AGGCAACTCT GTAAACTCCT CAGGGGGGCC AAAGCACTAA CAGACATAGT
3421 ACCACTAACT GAAGAAGCAG AATTAGAATT GGCAGAGAAC AGGGAAATTT TAAGAGAACC
3481 AGTACATGGA GTATATTATG ATCCATCAAA AGACTTGATA GCTGAAATAC AGAAACAGGG
3541 GCATGAACAA TGGACATATC AAATTTATCA AGAACCATTT AAAAATCTGA AAACAGGGAA
3601 GTATGCAAAA ATGAGGACTA CCCACACTAA TGATGTAAAA CAGTTAACAG AGGCAGTGCA
3661 AAAAATAGCC ATGGAAAGCA TAGTAATATG GGGAAAGACT CCTAAATTTA GACTACCCAT
3721 CCAAAAAGAA ACATGGGAGA CATGGTGGAC AGACTATTGG CAAGCCACCT GGATCCCTGA
3781 GTGGGAGTTT GTTAATACCC CTCCCCTAGT AAAATTATGG TACCAACTAG AAAAAGATCC
3841 CATAGCAGGA GTAGAAACTT TCTATGTAGA TGGAGCAACT AATAGGGAAG CTAAAATAGG
3901 AAAAGCAGGG TATGTTACTG ACAGAGGAAG GCAGAAAATT GTTACTCTAA CTAACACAAC
3961 AAATCAGAAG ACTGAGTTAC AAGCAATTCA GCTAGCTCTG CAGGATTCAG GATCAGAAGT
4021 AAACATAGTA ACAGACTCAC AGTATGCATT AGGAATCATT CAAGCACAAC CAGATAAGAG
4081 TGACTCAGAG ATATTTAACC AAATAATAGA ACAGTTAATA AACAAGGAAA GAATCTACCT
4141 GTCATGGGTA CCAGCACATA AAGGAATTGG GGGAAATGAA CAAGTAGATA AATTAGTAAG
4201 TAAGGGAATT AGGAAAGTGT TGTTTCTAGA TGGAATAGAT AAAGCTCAAG AAGAGCATGA
4261 AAGGTACCAC AGCAATTGGA GAGCAATGGC TAATGAGTTT AATCTGCCAC CCATAGTAGC
4321 AAAAGAAATA GTAGCTAGCT GTGATAAATG TCAGCTAAAA GGGGAAGCCA TACATGGACA
4381 AGTCGACTGT AGTCCAGGGA TATGGCAATT AGATTGTACC CATTTAGAGG GAAAAATCAT
4441 CCTGGTAGCA GTCCATGTAG CTAGTGGCTA CATGGAAGCA GAGGTTATCC CAGCAGAAAC
4501 AGGACAAGAA ACAGCATATT TTATATTAAA ATTAGCAGGA AGATGGCCAG TCAAAGTAAT
4561 ACATACAGAC AATGGCAGTA ATTTTACCAG TACTGCAGTT AAGGCAGCCT GTTGGTGGGC
4621 AGGTATCCAA CAGGAATTTG GAATTCCCTA CAATCCCCAA AGTCAGGGAG TGGTAGAATC
4681 CATGAATAAA GAATTAAAGA AAATAATAGG ACAAGTAAGA GATCAAGCTG AGCACCTTAA
4741 GACAGCAGTA CAAATGGCAG TATTCATTCA CAATTTTAAA AGAAAAGGGG GAATTGGGGG
4801 GTACAGTGCA GGGGAAAGAA TAATAGACAT AATAGCAACA GACATACAAA CTAAAGAATT
4861 ACAAAAACAA ATTATAAGAA TTCAAAATTT TCGGGTTTAT TACAGAGACA GCAGAGACCC
4921 TATTTGGAAA GGACCAGCCG AACTACTCTG GAAAGGTGAA GGGGTAGTAG TAATAGAAGA
4981 TAAAGGTGAC ATAAAGGTAG TACCAAGGAG GAAAGCAAAA ATCATTAGAG ATTATGGAAA
5041 ACAGATGGCA GGTGCTGATT GTGTGGCAGG TGGACAGGAT GAAGATTAGA GCATGGAATA
5101 GTTTAGTAAA GCACCATATG TATATATCAA GGAGAGCTAG TGGATGGGTC TACAGACATC
5161 ATTTTGAAAG CAGACATCCA AAAGTAAGTT CAGAAGTACA TATCCCATTA GGGGATGCTA
5221 GATTAGTAAT AAAAACATAT TGGGGTTTGC AGACAGGAGA AAGAGATTGG CATTTGGGTC
5281 ATGGAGTCTC CATAGAATGG AGACTGAGAG AATACAGCAC ACAAGTAGAC CCTGACCTGG
5341 CAGACCAGCT AATTCACATG CATTATTTTG ATTGTTTTAC AGAATCTGCC ATAAGACAAG
5401 CCATATTAGG ACACATAGTT TTTCCTAGGT GTGACTATCA AGCAGGACAT AAGAAGGTAG
5461 GATCTCTGCA ATACTTGGCA CTGACAGCAT TGATAAAACC AAAAAAGAGA AAGCCACCTC
5521 TGCCTAGTGT TAGAAAATTA GTAGAGGATA GATGGAACGA CCCCCAGAAG ACCAGGGGCC
5581 GCAGAGGGAA CCATACAATG AATGGACACT AGAGATTCTA GAAGAACTCA AGCAGGAAGC
5641 TGTCAGACAC TTTCCTAGAC CATGGCTCCA TAGCTTAGGA CAATATATCT ATGAAACCTA
5701 TGGGGATACT TGGACGGGAG TTGAAGCTAT AATAAGAGTA CTGCAACAAC TACTGTTCAT
5761 TCATTTCAGA ATTGGATGCC AACATAGCAG AATAGGCATC TTGCGACAGA GAAGAGCAAG
5821 AAATGGAGCC AGTAGATCCT AAACTAAAGC CCTGGAACCA TCCAGGAAGC CAACCTAAAA
5881 CAGCTTGTAA TAATTGCTTT TGCAAACACT GTAGCTATCA TTGTCTAGTT TGCTTTCAGA
```

FIGURE 11

```
5941 CAAAAGGTTT AGGCATTTCC TATGGCAGGA AGAAGCGGAG ACAGCGACGA AGCGCTCCTC
6001 CAAGTGGTGA AGATCATCAA AATCCTCTAT CAAAGCAGTA AGTACACATA GTAGATGTAA
6061 TGGTAAGTTT AAGTTTATTT AAAGGAGTAG ATTATAGATT AGGAGTAGGA GCATTGATAG
6121 TAGCACTAAT CATAGCAATA ATAGTGTGGA CCATAGCATA TATAGAATAT AGGAAATTGG
6181 TAAGACAAAA GAAAATAGAC TGGTTAATTA AAAGAATTAG GGAAAGAGCA GAAGACAGTG
6241 GCAATGAGAG TGATGGGGAC ACAGAAGAAT TGTCAACAAT GGTGGATATG GGGCATCTTA
6301 GGCTTCTGGA TGCTAATGAT TTGTAACACG GAGGACTTGT GGGTCACAGT CTACTATGGG
6361 GTACCTGTGT GGAGAGAAGC AAAAACTACT CTATTCTGTG CATCAGATGC TAAAGCATAT
6421 GAGACAGAAG TGCATAATGT CTGGGCTACA CATGCTTGTG TACCCACAGA CCCCAACCCA
6481 CAAGAAATAG TTTTGGGAAA TGTAACAGAA AATTTTAATA TGTGGAAAAA TAACATGGCA
6541 GATCAGATGC ATGAGGATAT AATCAGTTTA TGGGATCAAA GCCTAAAGCC ATGTGTAAAG
6601 TTGACCCCAC TCTGTGTCAC TTTAAACTGT ACAGATACAA ATGTTACAGG TAATAGAACT
6661 GTTACAGGTA ATACAAATGA TACCAATATT GCAAATGCTA CATATAAGTA TGAAGAAATG
6721 AAAAATTGCT CTTTCAATGC AACCACAGAA TTAAGAGATA AGAAACATAA AGAGTATGCA
6781 CTCTTTTATA AACTTGATAT AGTACCACTT AATGAAAATA GTAACAACTT TACATATAGA
6841 TTAATAAATT GCAATACCTC AACCATAACA CAAGCCTGTC CAAAGGTCTC TTTTGACCCG
6901 ATTCCTATAC ATTACTGTGC TCCAGCTGAT TATGCGATTC TAAAGTGTAA TAATAAGACA
6961 TTCAATGGGA CAGGACCATG TTATAATGTC AGCACAGTAC AATGTACACA TGGAATTAAG
7021 CCAGTGGTAT CAACTCAACT ACTGTTAAAT GGTAGTCTAG CAGAAGAAGG GATAATAATT
7081 AGATCTGAAA ATTTGACAGA GAATACCAAA ACAATAATAG TACATCTTAA TGAATCTGTA
7141 GAGATTAATT GTACAAGGCC CAACAATAAT ACAAGGAAAA GTGTAAGGAT AGGACCAGGA
7201 CAAGCATTCT ATGCAACAAA TGACGTAATA GGAAACATAA GACAAGCACA TTGTAACATT
7261 AGTACAGATA GATGGAATAA AACTTTACAA CAGGTAATGA AAAAATTAGG AGAGCATTTC
7321 CCTAATAAAA CAATAAAATT TGAACCACAT GCAGGAGGGG ATCTAGAAAT TACAATGCAT
7381 AGCTTTAATT GTAGAGGAGA ATTTTTCTAT TGCAATACAT CAAACCTGTT TAATAGTACA
7441 TACTACCCTA AGAATGGTAC ATACAAATAC AATGGTAATT CAAGCTTACC CATCACACTC
7501 CAATGCAAAA TAAAACAAAT TGTACGCATG TGGCAAGGGG TAGGACAAGC AATGTATGCC
7561 CCTCCCATTG CAGGAAACAT AACATGTAGA TCAAACATCA CAGGAATACT ATTGACACGT
7621 GATGGGGGAT TTAACAACAC AAACAACGAC ACAGAGGAGA CATTCAGACC TGGAGGAGGA
7681 GATATGAGGG ATAACTGGAG AAGTGAATTA TATAAATATA AAGTGGTAGA AATTAAGCCA
7741 TTGGGAATAG CACCCACTAA GGCAAAAAGA AGAGTGGTGC AGAGAAAAAA AAGAGCAGTG
7801 GGAATAGGAG CTGTGTTCCT TGGGTTCTTG GGAGCAGCAG GAAGCACTAT GGGCGCAGCG
7861 TCAATAACGC TGACGGTACA GGCCAGACAA CTGTTGTCTG GTATAGTGCA ACAGCAAAGC
7921 AATTTGCTGA AGGCTATAGA GGCGCAACAG CATATGTTGC AACTCACAGT CTGGGGCATT
7981 AAGCAGCTCC AGGCGAGAGT CCTGGCTATA GAAAGATACC TAAAGGATCA ACAGCTCCTA
8041 GGGATTTGGG GCTGCTCTGG AAGACTCATC TGCACCACTG CTGTGCCTTG GAACTCCAGT
8101 TGGAGTAATA AATCTGAAGC AGATATTTGG GATAACATGA CTTGGATGCA GTGGGATAGA
8161 GAAATTAATA ATTACACAGA AACAATATTC AGGTTGCTTG AAGACTCGCA AAACCAGCAG
8221 GAAAAGAATG AAAAAGATTT ATTAGAATTG GACAAGTGGA ATAATCTGTG GAATTGGTTT
8281 GACATATCAA ACTGGCTGTG GTATATAAAA ATATTCATAA TGATAGTAGG AGGCTTGATA
8341 GGTTTAAGAA TAATTTTTGC TGTGCTCTCT ATAGTGAATA GAGTTAGGCA GGGATACTCA
8401 CCTTTGTCAT TTCAGACCCT TACCCCAAGC CCGAGGGGAC TCGACAGGCT CGGAGGAATC
8461 GAAGAAGAAG GTGGAGAGCA AGACAGAGAC AGATCCATAC GATTGGTGAG CGGATTCTTG
8521 TCGCTTGCCT GGGACGATCT GCGGAGCCTG TGCCTCTTCA GCTACCACCG CTTGAGAGAC
8581 TTCATATTAA TTGCAGTGAG GGCAGTGGAA CTTCTGGGAC ACAGCAGTCT CAGGGGACTA
8641 CAGAGGGGGT GGGAGATCCT TAAGTATCTG GGAAGTCTTG TGCAGTATTG GGGTCTAGAG
8701 CTAAAAAAGA GTGCTATTAG TCCGCTTGAT ACCATAGCAA TAGCAGTAGC TGAAGGAACA
8761 GATAGGATTA TAGAATTGGT ACAAAGAATT TGTAGAGCTA TCCTCAACAT ACCTAGGAGA
8821 ATAAGACAGG GCTTTGAAGC AGCTTTGCTA TAAAATGGGA GGCAAGTGGT CAAAACGCAG
8881 CATAGTTGGA TGGCCTGCAG TAAGAGAAAG AATGAGAAGA ACTGAGCCAG CAGCAGAGGG
8941 AGTAGGAGCA GCGTCTCAAG ACTTAGATAG ACATGGGGCA CTTACAAGCA GCAACACACC
```

FIGURE 11

```
9001 TGCTACTAAT GAAGCTTGTG CCTGGCTGCA AGCACAAGAG GAGGACGGAG ATGTAGGCTT
9061 TCCAGTCAGA CCTCAGGTAC CTTTAAGACC AATGACTTAT AAGAGTGCAG TAGATCTCAG
9121 CTTCTTTTTA AAAGAAAAGG GGGGACTGGA AGGGTTAATT TACTCTAGGA AAAGGCAAGA
9181 AATCCTTGAT TTGTGGGTCT ATAACACACA AGGCTTCTTC CCTGATTGGC AAAACTACAC
9241 ATCGGGGCCA GGGGTCCGAT TCCCACTGAC CTTTGGATGG TGCTTCAAGC TAGTACCAGT
9301 TGACCCAAGG GAGGTGAAAG AGGCCAATGA AGGAGAAGAC AACTGTTTGC TACACCCTAT
9361 GAGCCAACAT GGAGCAGAGG ATGAAGATAG AGAAGTATTA AAGTGGAAGT TTGACAGCCT
9421 TCTAGCACAC AGACACATGG CCCGCGAGCT ACATCCGGAG TATTACAAAG ACTGCTGACA
9481 CAGAAGGGAC TTTCCGCCTG GGACTTTCCA CTGGGGCGTT CCGGGAGGTG TGGTCTGGGC
9541 GGGACTTGGG AGTGGTCACC CTCAGATGCT GCATATAAGC AGCTGCTTTT CGCTTGTACT
9601 GGGTCTCTCT CGGTAGACCA GATCTGAGCC TGGGAGCTCT CTGGCTATCT AGGGAACCCA
9661 CTGCTTAGGC CTCAATAAAG CTTGCCTTGA GTGCTCTAAG TAGTGTGTGC CCATCTGTTG
9721 TGTGACTCTG GTAACTAGAG ATCCCTCAGA CCCTTTGTGG TAGTGTGGAA AATCTCTAGC
9781 A
```

FIGURE 11

SEQ ID NO:34

GCTGAGGCAATGAGCCAAGCAACCAGCGCAAACATACTGATGCAGAGAAGCAATTT
CAAAGGCCCTAAAAGAATTATTAAATGTTTCAACTGTGGCAAGGAAGGGCACATAG
CTAGAAATTGTAGGGCCCCTAGGAAAAAAGGCTGTTGGAAATGTGGAAAGGAAGGA
CACCAAATGAAAGACTGTACTGAGAGGCAGGCTAA

FIGURE 12

975Pol wt until 6aa Int: (SEQ ID NO:35)
TTTTTTAGGGAAGATTTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTCCTTCAGAA
CAGAACAGAGCCAACAGCCCCACCAGCAGAGAGCTTCAAGTTCGAGGAGACAACCC
CCGCTCCGAAGCAGGAGCCGAAAGACAGGGAACCCTTAATTTCCCTCAAATCACTCT
TTGGCAGCGACCCCTTGTCTCAATAAAAGTAGGGGGTCAAATAAAGGAGGCTCTCTT
AGACACAGGAGCTGATGATACAGTATTAGAAGAAATGAGTTTGCCAGGAAAATGGA
AACCAAAAATGATAGGAGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAA
ATACTTATAGAAATTTGTGGAAAAAAGGCTATAGGTACAGTATTAATAGGACCTACA
CCTGTCAACATAATTGGAAGGAATATGTTGACTCAGCTTGGATGCACACTAAATTTT
CCAATTAGTCCCATTGAAACTGTGCCAGTAAAATTAAAGCCAGGAATGGATGGCCCA
AAGGTTAAACAATGGCCATTGACAGAAGAGAAAATAAAAGCATTAACAGCAATTTG
TGAAGAAATGGAGAAAGAAGGAAAAATTACAAAAATTGGGCCTGAAAATCCATATA
ACACTCCAGTATTTGCCATAAAAAAGAAGGACAGTACTAAGTGGAGAAAGTTAGTA
GATTTCAGGGAACTTAATAAAAGAACTCAAGACTTTTGGGAAGTTCAATTAGGAATA
CCACACCCAGCAGGGTTAAAAAAGAAAAAATCAGTGACAGTACTGGATGTGGGGGA
TGCATATTTTTCAGTTCCTTTAGATGAGGACTTCAGGAAATATACTGCATTCACCATA
CCTAGTATAAACAATGAAACACCAGGGATTAGATATCAATATAATGTGCTTCCACAG
GGATGGAAAGGATCACCATCAATATTCCAGAGTAGCATGACAAAAATCTTAGAGCC
CTTTAGAGCAAGAAATCCAGAAATAGTCATCTATCAATATATGGATGACTTGTATGT
AGGATCTGACTTAGAAATAGGGCAACATAGAGCAAAAATAGAGGAGTTAAGAAAAC
ATCTGTTAAGGTGGGGATTTACCACACCGGACAAGAAACATCAGAAAGAACCCCCA
TTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACAGTACAGCCTATAGAG
TTGCCAGAAAAGGAAAGCTGGACTGTCAATGATATACAGAAGTTAGTGGGAAAATT
AAATTGGGCCAGTCAGATTTACCCAGGAATTAAAGTAAGGCAACTTTGTAAACTCCT
TAGGGGGGCCAAAGCACTAACAGATATAGTACCACTAACTGAAGAAGCAGAATTAG
AATTGGCAGAGAACAGGGAAATTCTAAGAGAACCAGTACATGGAGTATATTATGAC
CCATCAAAAGACTTGGTAGCTGAAATACAGAAACAGGGGCATGACCAATGGACATA
TCAAATTTACCAAGAACCATTCAAAAACCTGAAAACAGGGAAGTATGCAAAAATGA
GGACTGCCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCT
ATGGAAAGCATAGTAATATGGGGAAAGACTCCTAAATTTAGACTACCCATCCAAAA
AGAAACATGGGAGACATGGTGGACAGACTATTGGCAAGCCACCTGGATTCCTGAGT
GGGAGTTTGTTAATACCCCTCCCTTAGTAAAATTATGGTACCAGCTAGAGAAAGAAC
CCATAATAGGAGCAGAAACTTTCTATGTAGATGGAGCAGCTAATAGGGAAACTAAA
ATAGGAAAAGCAGGGTATGTTACTGACAGAGGAAGGCAGAAAATTGTTTCTCTAAC
AGAAACAACAAATCAGAAGACTGAATTACAAGCAATTCAGCTAGCTTTGCAAGATTC
AGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAG
CACAACCAGATAAGAGTGAATCAGAGTTAGTCAACCAAATAATAGAACAATTAATA
AAAAAGGAAAAGGTCTACCTGTCATGGGTACCAGCACATAAAGGAATTGGAGGAAA
TGAACAAATAGATAAATTAGTAAGTAAGGGAATCAGGAAAGTGCTGTTTCTAGATG
GAATAGAT

FIGURE 13

SEQ ID NO:36

GGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCG
GC

FIGURE 14

SEQ ID NO: 37

GGIVIYQYMDDLYVGSGG

FIGURE 15

12_5/1ZA (SEQ ID NO:45)

TGGAAGGGTTAATTTACTCCAGGAAAAGGCAAGAGATCCTTGATTTATGGGTCTATC
ACACACAAGGCTACTTCCCTGATTGGCAAAACTACACACCGGGACCAGGGGTCAGA
TATCCACTGACCTTTGGATGGTGCTTCAAGCTAGTGCCAGTTGACCCAAGGGAAGTA
GAAGAGGCCAACGGAGGAGAAGACAACTGTTTGCTACACCCTATGAGCCAGTATGG
AATGGATGATGAACACAAAGAAGTGTTACAGTGGAAGTTTGACAGCAGCCTAGCAC
GCAGACACCTGGCCCGCGAGCTACATCCGGATTATTACAAAGACTGCTGACACAGA
AGGGACTTTCCGCCTGGGACTTTCCACTGGGGCGTTCCAGGGGGAGTGGTCTGGGCG
GGACTGGGAGTGGCCAGCCCTCAGATGCTGCATATAAGCAGCGGCTTTTCGCCTGTA
CTGGGTCTCTCTAGGTAGACCAGATCCGAGCCTGGGAGCTCTCTGTCTATCTGGGGA
ACCCACTGCTTAGGCCTCAATAAAGCTTGCCTTGAGTGCTCTAAGTAGTGTGTGCCC
ATCTGTTGTGTGACTCTGGTAACTCTGGTAACTAGAGATCCCTCAGACCCTTTGTGGT
AGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTGAG
ACCAGAGAAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGTGCACTCGGCAAGAG
GCGAGGGGGGCGACTGGTGAGTACGCCAAAATTTTTTTTGACTAGCGGAGGCTAGA
AGGAGAGAGATGGGTGCGAGAGCGTCAATATTAAGAGGGGGAAAATTAGACAAAT
GGGAAAAAATTAGGTTACGGCCAGGGGGGAGAAAACACTATATGCTAAAACACCTA
GTATGGGCAAGCAGAGAGCTGGAAAGATTTGCAGTTAACCCTGGCCTTTTAGAGAC
ATCAGACGGATGTAGAC AAATAATAAAACAGCTACAACCAGCTCTTCAGA
CAGGAACAGAGGAAATTAGATCATTATTTAACACAGTAGCAACTCTCTATTGTGTAC
ATAAAGGGATAGATGTACGAGACACCAAGGAAGCCTTAGACAAGATAGAGGAGGA
ACAAAACAAATGTCAGCAAAAAACACAGCAGGCGGAAGCGGCTGACAAAAAGGTC
AGTCAAAATTATCCTATAGTGCAGAACCTCCAAGGGCAAATGGTACACCAGGCCAT
ATCACCTAGAACCTTGAATGCATGGGTAAAAGTAATAGAGGAGAAGGCTTTTAGCC
CAGAGGTAATACCCATGTTTACAGCATTATCAGAAGGAGCCACCCCACAAGATTTA
AACACCATGTTAAATACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAG
ATACCATCAATGAGGAGGCTGCAGAATGGGATAGGTTACATCCAGTACATGCAGGG
CCTGTTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTA
CTAGTACCCTTCAAGAACAAATAGCATGGATGACAAGTAACCCACCTATCCCAGTA
GGGGACATCTATAAAAGGTGGATAATTCTGGGGTTAAATAAAATAGTAAGAATGTA
CAGCCCTGTCAGCATTTTAGACATAAAACAAGGACCAAAGGAACCCTTTAGAGACT
ATGTAGACCGGTTCTTCAAAACTTTAAGAGCTGAACAATCTACACAAGAGGTAAAA
AATTGGATGACAGACACCTTGTTAGTCCAAAATGCGAACCCAGATTGTAAGACCATT
TTAAGAGCATTAGGACCAGGGGCTTCATTAGAAGAAATGATGACAGCATGTCAGGG
AGTGGGAGGACCTAGCCACAAAGCAAGAGTTTTGGCTGAGGCAATGAGCCAAGCAA
ACAATACAAGTGTAATGATACAGAAAAGCAATTTTAAAGGCCCTAGAAGAGCTGTT
AAATGTTTCAACTGTGGCAGGGAAGGGCACATAGCCAGGAATTGCAGGGCCCCTAG
GAAAAGGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGACTGTACT
GAGAGGCAGGCTAATTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGGAGGCCAGG
GAATTTCCTTCAGAGCAGACCAGAGCCAACAGCCCCACCACTAGAACCAACAGCCC
CACCAGCAGAGAGCTTCAAGTTCAAGGAGACTCCGAAGCAGGAGCCGAAAGACAG
GGAACCTTTAACTTCCCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAA

FIGURE 16

GTAGCGGGCCAAACAAAGGAGGCTCTTTTAGATACAGGAGCAGATGATACAGTACT
AGAAGAAATAAACTTGCCAGGAAAATGGAAACCAAAAATGATAGGAGGAATTGGA
GGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAGG
GCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTG
TTGACTCAGCTTGGATGCACACTAAATTTTCCAATTAGCCCCATTGAAACTGTACCA
GTAAAATTAAAGCCAGGAATGGATGGCCCAAAGGTTAAACAATGGCCATTGACAGA
AGAAAAAATAAAAGCATTAACAGAAATTTGTGAGGAAATGGAGAAGGAAGGAAAA
ATTACAAAAATTGGGCCTGAAAATCCATATAACACTCCAGTATTTGCCATAAAGAAG
AAGGACAGTACAAAGTGGAGAAAATTAGTAGATTTCAGGGAACTCAATAAAAGAAC
TCAAGACTTTTGGGAAGTCCAATTAGGAATACCACACCCAGCAGGGTTAAAAAAGA
AAAAATCAGTGACAGTACTGGATGTGGGAGATGCATATTTTTCAGTCCCTTTAGATG
AGAGCTTCAGAAAATATACTGCATTCACCATACCTAGTATAAACAATGAAACACCA
GGGATTAGATATCAATATAATGTTCTTCCACAGGGATGGAAAGGATCACCAGCAA
TATTCCAGAGTAGCATGACAAGAATCTTAGAGCCCTTTAGAACACAAAACCCAGAA
GTAGTTATCTATCAATATATGGATGACTTATATGTAGGATCTGACTTAGAAATAGGG
CAACATAGAGCAAAAATAGAGGAGTTAAGAGGACACCTATTGAAATGGGGATTTAC
CACACCAGACAAGAAACATCAGAAAGAACCCCCATTTCTTTGGATGGGGTATGAAC
TCCATCCTGACAAATGGACAGTACAGCCTATACAGCTGCCAGAAAAGGAGAGCTGG
ACTGTCAATGATATACAGAAGTTAGTGGGAAAGTTAAACTGGGCAAGTCAGATTTA
CCCAGGGATTAAAGTAAGGCAACTGTGTAAACTCCTTAGGGGAGCCAAAGCACTAA
CAGACATAGTGCCACTGACTGAAGAAGCAGAATTAGAATTGGCTGAGAACAGGGA
AATTCTAAAAGAACCAGTACATGGAGTATATTATGACCCATCAAAAGATTTAATAG
CTGAAATACAGAAACAGGGGAATGACCAATGGACATATCAAATTTACCAAGAACC
ATTTAAAAATCTGAGAACAGGAAAGTATGCAAAAATGAGGACTGCCCACACTAATG
ATGTGAAACAGTTAGCAGAGGCAGTGCAAAAGATAACCCAGGAAAGCATAGTAATA
TGGGGAAAAACTCCTAAATTTAGACTACCCATCCCAAAAGAAACATGGGAGACATG
GTGGTCAGACTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTCAATACCCC
TCCCCTAGTAAAATTGTGGTACCAGCTGGAAAAAGAACCCATAGTAGGGGCAGAAA
CTTTCTATGTAGATGGAGCAGCCAATAGGGAAACTAAAATAGGAAAAGCAGGGTAT
GTCACTGACAAAGGAAGGCAGAAAGTTGTTTCCTTCACTGAAACAACAAATCAGAA
GACTGAATTACAAGCAATTCAGCTAGCTTTGCAGGATTCAGGGCCAGAAGTAAACA
TAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAACCAGATAAGAGT
GAATCAGAATTAGTCAGTCAAATAATAGAACAGTTGATAAAAAAGGAAAAAGTCTA
CCTATCATGGGTACCAGCACATAAAGGAATTGGAGGAAATGAACAAGTAGACAAAT
TAGTAAGTAGTGGAATCAGAAAAGTACTGTTTCTAGATGGAATAGATAAAGCTCAA
GAAGAGCATGAAAAATATCACAGCAATTGGAGAGCAATGGCTAGTGAGTTTAATCT
GCCACCCATAGTAGCAAAGGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAG
GGGAAGCCATGCATGGACAAGTCGACTGTAGTCCAGGAATATGGCAATTAGACTGT
ACACATTTAGAAGGAAAAATCATCCTAGTAGCAGTCCATGTAGCCAGTGGCTACAT
GGAAGCAGAGGTTATCCCAGCAGAAACAGGACAAGAAACAGCATACTTTATACTAA
AATTAGCAGGAAGATGGCCAGTCAAAGTAATACATACAGATAATGGCAGTAATTTC
ACCAGTACCGCAGTTAAGGCAGCCTGTTGGTGGGCAGATATCCAACGGGAATTTGG
AATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCCATGAATAAAGAATTAA

FIGURE 16

AGAAAATCATAGGGCAAGTAAGAGATCAAGCTGAGCACCTTAAGACAGCAGTACAA
ATGGCAGTATTCATTCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGC
AGGGGAGAGAATAATAGACATAATAGCATCAGACATACAAACTAAAGAATTACAAA
AACAAATTATAAAAATTCAAAATTTTCGGGTTTATTACAGAGACAGCAGAGACCCTA
TTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACAA
GATAATAGTGATATAAAGGTAGTACCAAGAAGGAAAGCAAAAATCATTAAGGACTA
TGGAAAACAGATGGCAGGTGCTGATTGTGTGGCAGGTAGACAGGATGAAGATTAGA
ACATGGCACAGTTTAGTAAAGCACCATATGTATGTTTCGAGGAGAGCTGATGGATGG
TTCTACAGACATCATTATGAAAGCAGACACCCAAAAGTAAGTTCAGAAGTACACAT
CCCATTAGGAGATGCCAGGTTAGTAATAAAAACATATTGGGGTCTGCAGACAGGAG
AAAGAGCTTGGCATTTGGGTCACGGAGTCTCCATAGAATGGAGATTGAGAAGATAT
AGCACACAAGTAGACCCTGACCTGACAGACCAACTAATTCATATGCATTATTTTGAT
TGTTTTGCAGAATCTGCCATAAGGAAAGCCATACTAGGACAGATAGTTAGCCCTAA
GTGTGACTATCAAGCAGGACATAACAAGGTAGGATCTCTACAATACTTGGCACTGA
CAGCATTGATAAAACCAAAAAAGATAAAGCCACCTCTGCCTAGTGTTAGGAAATTA
GTAGAGGATAGATGGAACAAGCCCCAGAAGACCAGGGGCCGCAGAGGGAACCATA
CAATGAATGGACACTAGAGCTTTTAGAAGAACTCAAGCAGGAAGCTGTCAGACACT
TTCCTAGACCATGGCTCCATAACTTAGGACAACATATCTATGAAACCTATGGAGATA
CTTGGACAGGAGTTGAAGCAATAATAAGAATCCTGCAACAATTACTGTTTATTCATT
TCAGGATTGGGTGCCATCATAGCAGAATAGGCATTTTGCGACAGAGAAGAGCAAGA
AATGGAGCCAATAGATCCTAACCTAGAACCCTGGAACCATCCAGGAAGTCAGCCTA
AAACTGCTTGTAATGGGTGTTACTGTAAACGTTGCAGCTATCATTGTCTAGTTTGCTT
TCAGAAAAAAGGCTTAGGCATTTACTATGGCAGGAAGAAGCGGAGACAGCGACGAA
GCGCTCCTCCAAGCAATAAAGATCATCAAGATCCTCTACCAAAGCAGTAAGTACCG
AATAGTATATGTAATGTTAGATTTAACTGCAAGAATAGATTCTAGATTAGGAATAGG
AGCATTGATAGTAGCACTAATCATAGCAATAATAGTGTGGACCATAGTATATATAG
AATATAGGAAATTGGTAAGGCAAAGGAAAATAGACTGGTTAGTTAAAAGGATTAGG
GAAAGAGCAGAAGACAGTGGCAATGAGAGCGAGGGGGATACTGAAGAATTATCGA
CACTGGTGGATATGGGGCATCTTAGGCTTTTGGATGCTAATGATGTGTAATGTGAA
GGGCTTGTGGGTCACAGTCTACTACGGGGTACCTGTGGGGAGAGAAGCAAAAACT
ACTCTATTTTGTGCATCAGATGCTAAAGCATATGAGAAAGAAGTGCATAATGTCTG
GGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTGATTTTGGGC
AATGTAACAGAAAATTTTAACATGTGGAAAAATGACATGGTGGATCAGATGCAGG
AAGATATAATCAGTTTATGGGATCAAAGCCTTAAGCCATGTGTAAAATTGACCCCA
CTCTGTGTCACTTTAAACTGTACAAATGCAACTGTTAACTACAATAATACCTCTAAA
GACATGAAAAATTGCTCTTTCTATGTAACCACAGAATTAAGAGATAAGAAAAAGAA
AGAAAATGCACTTTTTTATAGACTTGATATAGTACCACTTAATAATAGGAAGAATGG
GAATATTAACAACTATAGATTAATAAATTGTAATACCTCAGCCATAACACAAGCCTG
TCCAAAAGTCTCGTTTGACCCAATTCCTATACATTATTGTGCTCCAGCTGGTTATGCG
CCTCTAAAATGTAATAATAAGAAATTCAATGGAATAGGACCATGCGATAATGTCAG
CACAGTACAATGTACACATGGAATTAAGCCAGTGGTATCAACTCAATTACTGTTAAA
TGGTAGCCTAGCAGAAGAAGAGATAATAATTAGATCTGAAAATCTGACAAACAATG
TCAAAACAATAATAGTACATCTTAATGAATCTATAGAGATTAAATGTACAAGACC

FIGURE 16

TGGCAATAATACAAGAAAGAGTGTGAGAATAGGACCAGGACAAGCATTCTATGCA
ACAGGAGACATAATAGGAGATATAAGACAAGCACATTGTAACATTAGTAAAAATGA
ATGGAATACAACTTTACAAAGGGTAAGTCAAAAATTACAAGAACTCTTCCCTAATA
GTACAGGGATAAAATTTGCACCACACTCAGGAGGGGACCTAGAAATTACTACACAT
AGCTTTAATTGTGGAGGAGAATTTTTCTATTGCAATACAACAGACCTGTTTAATAGT
ACATACAGTAATGGTACATGCACTAATGGTACATGCATGTCTAATAATACAGAGCG
CATCACACTCCAATGCAGAATAAAACAAATTATAAACATGTGGCAGGAGGTAGGAC
GAGCAATGTATGCCCCTCCCATTGCAGGAAACATAACATGTAGATCAAATATTACA
GGACTACTATTAACACGTGATGGAGGAGATAATAATACTGAAACAGAGACATTCAG
ACCTGGAGGAGGAGACATGAGGGACAATTGGAGAAGTGAATTATATAAATACAAG
GTGGTAGAAATTAAACCATTAGGAGTAGCACCCACTGCTGCAAAAAGGAGAGTGGT
GGAGAGAGAAAAAAGAGCAGTAGGAATAGGAGCTGTGTTCCTTGGGTTCTTGGGAG
CAGCAGGAAGCACTATGGGCGCAGCATCAATAACGCTGACGGTACAGGCCAGACAA
TTATTGTCTGGTATAGTGCAACAGCAAAGTAATTTGCTGAGGGCTATAGAGGCGCAA
CAGCATATGTTGCAACTCACGGTCTGGGGCATTAAGCAGCTCCAGGCAAGAGTCCTG
GCTATAGAGAGATACCTACAGGATCAACAGCTCCTAGGACTGTGGGGCTGCTCTGG
AAAACTCATCTGCACCACTAATGTGCTTTGGAACTCTAGTTGGAGTAATAAAACTCA
AAGTGATATTTGGGATAACATGACCTGGATGCAGTGGGATAGGGAAATTAGTAATT
ACACAAACACAATATACAGGTTGCTTGAAGACTCGCAAAGCCAGCAGGAAAGAAA
TGAAAAAGATTTACTAGCATTGGACAGGTGGAACAATCTGTGGAATTGGTTTAGCAT
AACAAATTGGCTGTGGTATATAAAAATATTCATAATGATAGTAGGAGGCTTGATAG
GTTTAAGAATAATTTTTGCTGTGCTCTCTCTAGTAAATAGAGTTAGGCAGGGATACT
CACCCTTGTCATTGCAGACCCTTATCCCAAACCCGAGGGGACCCGACAGGCTCGGA
GGAATCGAAGAAGAAGGTGGAGAGCAAGACAGCAGCAGATCCATTCGATTAGTGA
GCGGATTCTTGACACTTGCCTGGGACGACCTACGAAGCCTGTGCCTCTTCTGCTACC
ACCGATTGAGAGACTTCATATTAATTGTAGTGAGAGCAGTGGAACTTCTGGGACAC
AGTAGTCTCAGGGGACTGCAGAGGGGGTGGGGAACCCTTAAGTATTTGGGGAGTCT
TGTGCAATATTGGGGTCTAGAGTTAAAAAAGAGTGCTATTAATCTGCTTGATACTAT
AGCAATAGCAGTAGCTGAAGGAACAGATAGGATTCTAGAATTCATACAAAACCTTT
GTAGAGGTATCCGCAACGTACCTAGAAGAATAAGACAGGGCTTCGAAGCAGCTTTG
CAATAAAATGGGGGGCAAGTGGTCAAAAAGCAGTATAATTGGATGGCCTGAAGTAA
GAGAAAGAATCAGACGAACTAGGTCAGCAGCAGAGGGAGTAGGATCAGCGTCTCA
AGACTTAGAGAAACATGGGGCACTTACAACCAGCAACACAGCCCACAACAATGCTG
CTTGCGCCTGGCTGGAAGCGCAAGAGGAGGAAGGAGAAGTAGGCTTTCCAGTCAGA
CCTCAGGTACCTTTAAGACCAATGACTTATAAAGCAGCAATAGATCTCAGCTTCTTT
TTAAAAGAAAAGGGGGGACTGGAAGGGTTAATTTACTCCAAGAAAAGGCAAGAGAT
CCTTGATTTGTGGGTTTATAACACACAAGGCTTCTTCCCTGATTGGCAAAACTACAC
ACCGGGACCAGGGGTCAGATTTCCACTGACCTTTGGATGGTACTTCAAGCTAGAGCC
AGTCGATCCAAGGGAAGTAGAAGAGGCCAATGAAGGAGAAAACAACTGTTTACTAC
ACCCTATGAGCCAGCATGGAATGGAGGATGAAGACAGAGAAGTATTAAGATGGAAG
TTTGACAGTACGCTAGCACGCAGACACATGGCCCGCGAGCTACATCCGGAGTATTAC
AAAGACTGCTGACACAGAAGGGACTTTCCGCTGGGACTTTCCACTGGGGCGTTCCAG
GAGGTGTGGTCTGGGCGGGACAGGGGAGTGGTCAGCCCTGAGATGCTGCATATAAG
CAGCTGCTTTTCGCCTGTACTGGGTCTCTCTAGGTAGACCAGATCTGAGCCCGGGAG

FIGURE 16

CTCTCTGGCTATCTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTG
CCTTGAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGA
CCACTTGTGGTAGTGTGGAAAATCTCTAGCA

FIGURE 16

>C4_Env_TV1_C_ZA_opt_short (SEQ ID NO:46)

CATCACCCTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGGGCCAGGCCATGTACGCCCCCCCCATCG
CCGGCAACATCACCTGC

FIGURE 17

>C4_Env_TV1_C_ZA_opt (SEQ ID NO:47)

CTGCCCATCACCCTGCAGTGCAAGATCAAGCAGATCGTGCGCATGTGGCAGGGCGTGGGCCAGGCCATGTACGCCCCCCC
CATCGCCGGCAACATCACCTGCCGCAGCAACATCACCGGCATCCTGCTGACCCGCGACGGCGGC

FIGURE 18

>C4_Env_TV1_C_ZA_wt (SEQ ID NO:48)

TTACCCATCACACTCCAATGCAAAATAAAACAAATTGTACGCATGTGGCAAGGGGTAGGACAAGCAATGTATGCCCCTCC
CATTGCAGGAAACATAACATGTAGATCAAACATCACAGGAATACTATTGACACGTGATGGGGGA

FIGURE 19

>Envgp160_TV1_C_ZAopt (SEQ ID NO:49)

ATGCGCGTGATGGGCACCCAGAAGAACTGCCAGCAGTGGTGGATCTGGGGCATCCTGGGCTTCTGGATGCTGATGATCTG
CAACACCGAGGACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGCGCGAGGCCAAGACCACCCTGTTCTGCGCCA
GCGACGCCAAGGCCTACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCCACCGACCCCAACCCCCAG
GAGATCGTGCTGGGCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGCCGACCAGATGCACGAGGACATCAT
CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCGACACCAACG
TGACCGGCAACCGCACCGTGACCGGCAACACCAACGACACCAACATCGCCAACGCCACCTACAAGTACGAGGAGATGAAG
AACTGCAGCTTCAACGCCACCACCGAGCTGCGCGACAAGAAGCACAAGGAGTACGCCCTGTTCTACAAGCTGGACATCGT
GCCCCTGAACGAGAACAGCAACAACTTCACCTACCGCCTGATCAACTGCAACACCAGCACCATCACCCAGGCCTGCCCCA
AGGTGAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCCGCCGACTACGCCATCCTGAAGTGCAACAACAAGACCTTC
AACGGCACCGGCCCCTGCTACAACGTGAGCACCGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGAGCACCCAGCTGCT
GCTGAACGGCAGCCTGGCCGAGGAGGGCATCATCATCCGCAGCGAGAACCTGACCGAGAACACCAAGACCATCATCGTGC
ACCTGAACGAGAGCGTGGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCGTGCGCATCGGCCCCGGCCAG
GCCTTCTACGCCACCAACGACGTGATCGGCAACATCCGCCAGGCCCACTGCAACATCAGCACCGACCGCTGGAACAAGAC
CCTGCAGCAGGTGATGAAGAAGCTGGGCGAGCACTTCCCCAACAAGACCATCAAGTTCGAGCCCCACGCCGGCGGCGACC
TGGAGATCACCATGCACAGCTTCAACTGCCGCGGCGAGTTCTTCTACTGCAACACCAGCAACCTGTTCAACAGCACCTAC
TACCCCAAGAACGGCACCTACAAGTACAACGGCAACAGCAGCCTGCCCATCACCCTGCAGTGCAAGATCAAGCAGATCGT
GCGCATGTGGCAGGGCGTGGGCCAGGCCATGTACGCCCCCCCCATCGCCGGCAACATCACCTGCCGCAGCAACATCACCG
GCATCCTGCTGACCCGCGACGGCGGCTTCAACAACACCAACAACGACACCGAGGAGACCTTCCGCCCCGGCGGCGGCGAC
ATGCGCGACAACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGGAGATCAAGCCCCTGGGCATCGCCCCCACCAAGGC
CAAGCGCCGCGTGGTGCAGCGCAAGAAGCGCGCCGTGGGCATCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCA
GCACCATGGGCGCCGCCAGCATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAGAGCAAC
CTGCTGAAGGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCGTGCT
GGCCATCGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCCGCCTGATCTGCACCACCGCCG
TGCCCTGGAACAGCAGCTGGAGCAACAAGAGCGAGGCCGACATCTGGGACAACATGACCTGGATGCAGTGGGACCGCGAG
ATCAACAACTACACCGAGACCATCTTCCGCCTGCTGGAGGACAGCCAGAACCAGCAGGAGAAGAACGAGAAGGACCTGCT
GGAGCTGGACAAGTGGAACAACCTGTGGAACTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGA
TCGTGGGCGGCCTGATCGGCCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCC
CTGAGCTTCCAGACCCTGACCCCCAGCCCCCGCGGCCTGGACCGCCTGGGCGGCATCGAGGAGGAGGGCGGCGAGCAGGA
CCGCGACCGCAGCATCCGCCTGGTGAGCGGCTTCCTGAGCCTGGCCTGGGACGACCTGCGCAGCCTGTGCCTGTTCAGCT
ACCACCGCCTGCGCGACTTCATCCTGATCGCCGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGCCTGCGCGGCCTGCAG
CGCGGCTGGGAGATCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGCCTGGAGCTGAAGAAGAGCGCCATCAGCCC
CCTGGACACCATCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGCTGGTGCAGCGCATCTGCCGCGCCATCC
TGAACATCCCCCGCCGCATCCGCCAGGGCTTCGAGGCCGCCCTGCTGTAA

FIGURE 20

>Envgp160_TV1_C_ZAwt (SEQ ID NO:50)

ATGAGAGTGATGGGGACACAGAAGAATTGTCAACAATGGTGGATATGGGGCATCTTAGGCTTCTGGATGCTAATGATTTG
TAACACGGAGGACTTGTGGGTCACAGTCTACTATGGGGTACCTGTGTGGAGAGAAGCAAAAACTACTCTATTCTGTGCAT
CAGATGCTAAAGCATATGAGACAGAAGTGCATAATGTCTGGGCTACACATGCTTGTGTACCCACAGACCCCAACCCACAA
GAAATAGTTTTGGGAAATGTAACAGAAAATTTTAATATGTGGAAAAATAACATGGCAGATCAGATGCATGAGGATATAAT
CAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAGTTGACCCCACTCTGTGTCACTTTAAACTGTACAGATACAAATG
TTACAGGTAATAGAACTGTTACAGGTAATACAAATGATACCAATATTGCAAATGCTACATATAAGTATGAAGAAATGAAA
AATTGCTCTTTCAATGCAACCACAGAATTAAGAGATAAGAAACATAAAGAGTATGCACTCTTTTATAAACTTGATATAGT
ACCACTTAATGAAAATAGTAACAACTTTACATATAGATTAATAAATTGCAATACCTCAACCATAACACAAGCCTGTCCAA
AGGTCTCTTTTGACCCGATTCCTATACATTACTGTGCTCCAGCTGATTATGCGATTCTAAAGTGTAATAATAAGACATTC
AATGGGACAGGACCATGTTATAATGTCAGCACAGTACAATGTACACATGGAATTAAGCCAGTGGTATCAACTCAACTACT
GTTAAATGGTAGTCTAGCAGAAGAAGGGATAATAATTAGATCTGAAAATTTGACAGAGAATACCAAAACAATAATAGTAC
ATCTTAATGAATCTGTAGAGATTAATTGTACAAGGCCCAACAATAATACAAGGAAAAGTGTAAGGATAGGACCAGGACAA
GCATTCTATGCAACAAATGACGTAATAGGAAACATAAGACAAGCACATTGTAACATTAGTACAGATAGATGGAATAAAAC
TTTACAACAGGTAATGAAAAAATTAGGAGAGCATTTCCCTAATAAAACAATAAAATTTGAACCACATGCAGGAGGGGATC
TAGAAATTACAATGCATAGCTTTAATTGTAGAGGAGAATTTTTCTATTGCAATACATCAAACCTGTTTAATAGTACATAC
TACCCTAAGAATGGTACATACAAATACAATGGTAATTCAAGCTTACCCATCACACTCCAATGCAAAATAAAACAAATTGT
ACGCATGTGGCAAGGGGTAGGACAAGCAATGTATGCCCCTCCCATTGCAGGAAACATAACATGTAGATCAAACATCACAG
GAATACTATTGACACGTGATGGGGGATTTAACAACACAAACAACGACACAGAGGAGACATTCAGACCTGGAGGAGGAGAT
ATGAGGGATAACTGGAGAAGTGAATTATATAAATATAAAGTGGTAGAAATTAAGCCATTGGGAATAGCACCCACTAAGGC
AAAAAGAAGAGTGGTGCAGAGAAAAAAAAGAGCAGTGGGAATAGGAGCTGTGTTCCTTGGGTTCTTGGGAGCAGCAGGAA
GCACTATGGGCGCAGCGTCAATAACGCTGACGGTACAGGCCAGACAACTGTTGTCTGGTATAGTGCAACAGCAAAGCAAT
TTGCTGAAGGCTATAGAGGCGCAACAGCATATGTTGCAACTCACAGTCTGGGGCATTAAGCAGCTCCAGGCGAGAGTCCT
GGCTATAGAAAGATACCTAAAGGATCAACAGCTCCTAGGGATTTGGGGCTGCTCTGGAAGACTCATCTGCACCACTGCTG
TGCCTTGGAACTCCAGTTGGAGTAATAAATCTGAAGCAGATATTTGGGATAACATGACTTGGATGCAGTGGGATAGAGAA
ATTAATAATTACACAGAAACAATATTCAGGTTGCTTGAAGACTCGCAAAACCAGCAGGAAAAGAATGAAAAAGATTTATT
AGAATTGGACAAGTGGAATAATCTGTGGAATTGGTTTGACATATCAAACTGGCTGTGGTATATAAAAATATTCATAATGA
TAGTAGGAGGCTTGATAGGTTTAAGAATAATTTTTGCTGTGCTCTCTATAGTGAATAGAGTTAGGCAGGGATACTCACCT
TTGTCATTTCAGACCCTTACCCCAAGCCCGAGGGGACTCGACAGGCTCGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGA
CAGAGACAGATCCATACGATTGGTGAGCGGATTCTTGTCGCTTGCCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCT
ACCACCGCTTGAGAGACTTCATATTAATTGCAGTGAGGGCAGTGGAACTTCTGGGACACAGCAGTCTCAGGGGACTACAG
AGGGGGTGGGAGATCCTTAAGTATCTGGGAAGTCTTGTGCAGTATTGGGGTCTAGAGCTAAAAAAGAGTGCTATTAGTCC
GCTTGATACCATAGCAATAGCAGTAGCTGAAGGAACAGATAGGATTATAGAATTGGTACAAAGAATTTGTAGAGCTATCC
TCAACATACCTAGGAGAATAAGACAGGGCTTTGAAGCAGCTTTGCTATAA

FIGURE 21

>Gag_TV1_C_ZAopt (SEQ ID NO:51)

ATGGGCGCCCGCGCCAGCATCCTGAGCGGCGGCAAGCTGGACAAGTGGGAGCGCATCCGCCTGCGCCCCGGCGGCAAGAA
GCACTACATGCTGAAGCACCTGGTGTGGGCCAGCCGCGAGCTGGAGCGCTTCGCCCTGAACCCCGGCCTGCTGGAGACCA
GCGAGGGCTGCAAGCAGATCATCAAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGGAGCTGCGCAGCCTGTTCAAC
ACCGTGGCCACCCTGTACTGCGTGCACAAGGGCATCGAGGTGCGCGACACCAAGGAGGCCCTGGACAAGATCGAGGAGGA
GCAGAACAAGTGCCAGCAGAAGGCCCAGCAGGCCAAGGCCGCCGACGAGAAGGTGAGCCAGAACTACCCCATCGTGCAGA
ACGCCCAGGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCCTGAACGCCTGGATCAAGGTGATCGAGGAGAAGGCC
TTCAACCCCGAGGAGATCCCCATGTTCACCGCCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAACACCATGCTGAACAC
CGTGGGCGGCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAACGAGGAGGCCGCCGAGTGGGACCGCACCCACC
CCGTGCACGCCGGCCCCGTGGCCCCCGGCCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCACCACCAGCACCCTG
CAGGAGCAGATCGCCTGGATGACCAGCAACCCCCCCATCCCCGTGGAGGACATCTACAAGCGCTGGATCATCCTGGGCCT
GAACAAGATCGTGCGCATGTACAGCCCCGTGAGCATCCTGGACATCAAGCAGGGCCCCAAGGAGCCCTTCCGCGACTACG
TGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAGGCCACCCAGGACGTGAAGAACTGGATGACCGACACCCTGCTGGTG
CAGAACGCCAACCCCGACTGCAAGACCATCCTGCGCGCCCTGGGCCCCGGCGCCAGCCTGGAGGAGATGATGACCGCCTG
CCAGGGCGTGGGCGGCCCCAGCCACAAGGCCCGCGTGCTGGCCGAGGCCATGAGCCAGGCCAACAGCAACATCCTGGTGC
AGCGCAGCAACTTCAAGGGCAGCAACCGCATCATCAAGTGCTTCAACTGCGGCAAGGTGGGCCACATCGCCCGCAACTGC
CGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCCAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAA
CTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCCGGCAACTTCCTGCAGAACCGCCCCGAGCCCACCGCCCCCC
CCGCCGAGCCCACCGCCCCCCCCGCCGAGAGCTTCCGCTTCGAGGAGACCACCCCCGTGCCCCGCAAGGAGAAGGAGCGC
GAGCCCCTGACCAGCCTGAAGAGCCTGTTCGGCAGCGACCCCCTGAGCCAGTAA

FIGURE 22

>Gag_TV1_C_ZAwt (SEQ ID NO:52)

ATGGGTGCGAGAGCGTCAATATTAAGCGGCGGAAAATTAGATAAATGGGAAAGAATTAGGTTAAGGCCAGGGGGAAAGAA
ACATTATATGTTAAAACATCTAGTATGGGCAAGCAGGGAGCTGGAAAGATTTGCACTTAACCCTGGCCTGTTAGAAACAT
CAGAAGGCTGTAAACAAATAATAAAACAGCTACAACCAGCTCTTCAGACAGGAACAGAGGAACTTAGATCATTATTCAAC
ACAGTAGCAACTCTCTATTGTGTACATAAAGGGATAGAGGTACGAGACACCAAGGAAGCCTTAGACAAGATAGAGGAAGA
ACAAAACAAATGTCAGCAAAAAGCACAACAGGCAAAAGCAGCTGACGAAAAGGTCAGTCAAAATTATCCTATAGTACAGA
ATGCCCAAGGGCAAATGGTACACCAAGCTATATCACCTAGAACATTGAATGCATGGATAAAAGTAATAGAGGAAAAGGCT
TTCAATCCAGAGGAAATACCCATGTTTACAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACAATGTTAAATAC
AGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGATACCATCAATGAGGAGGCTGCAGAATGGGATAGGACACATC
CAGTACATGCAGGGCCTGTTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTT
CAGGAACAAATAGCATGGATGACAAGTAATCCACCTATTCCAGTAGAAGACATCTATAAAAGATGGATAATTCTGGGGTT
AAATAAAATAGTAAGAATGTATAGCCCTGTTAGCATTTTGGACATAAAACAAGGGCCAAAAGAACCCTTTAGAGACTATG
TAGACCGGTTCTTTAAAACCTTAAGAGCTGAACAAGCTACACAAGATGTAAAGAATTGGATGACAGACACCTTGTTGGTC
CAAAATGCGAACCCAGATTGTAAGACCATTTTAAGAGCATTAGGACCAGGGGCCTCATTAGAAGAAATGATGACAGCATG
TCAGGGAGTGGGAGGACCTAGCCATAAAGCAAGAGTGTTGGCTGAGGCAATGAGCCAAGCAAACAGTAACATACTAGTGC
AGAGAAGCAATTTTAAAGGCTCTAACAGAATTATTAAATGTTTCAACTGTGGCAAAGTAGGGCACATAGCCAGAAATTGC
AGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGACAGGAAGGACACCAAATGAAAGACTGTACTGAGAGGCAGGCTAA
TTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGGAGGCCAGGGAATTTCCTCCAGAACAGACCAGAGCCAACAGCCCCAC
CAGCAGAACCAACAGCCCCACCAGCAGAGAGCTTCAGGTTCGAGGAGACAACCCCCGTGCCGAGGAAGGAGAAAGAGAGG
GAACCTTTAACTTCCCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAA

FIGURE 23

>Gag_TV1_ZA_MHRopt (SEQ ID NO:53)

GACATCAAGCAGGGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAGACC

FIGURE 24

>Gag_TV1_ZA_MHRwt (SEQ ID NO:54)

GACATAAAACAAGGGCCAAAAGAACCCTTTAGAGACTATGTAGACCGGTTCTTTAAAACC

FIGURE 25

>Nef_TV1_C_ZAopt (SEQ ID NO:55)

ATGGGCGGCAAGTGGAGCAAGCGCAGCATCGTGGGCTGGCCCGCCGTGCGCGAGCGCATGCGCCGCACCGAGCCCGCCGC
CGAGGGCGTGGGCGCCGCCAGCCAGGACCTGGACCGCCACGGCGCCCTGACCAGCAGCAACACCCCCGCCACCAACGAGG
CCTGCGCCTGGCTGCAGGCCCAGGAGGAGGACGGCGACGTGGGCTTCCCCGTGCGCCCCCAGGTGCCCCTGCGCCCCATG
ACCTACAAGAGCGCCGTGGACCTGAGCTTCTTCCTGAAGGAGAAGGGCGGCCTGGAGGGCCTGATCTACAGCCGCAAGCG
CCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGCTTCTTCCCCGACTGGCAGAACTACACCAGCGGCCCCGGCG
TGCGCTTCCCCCTGACCTTCGGCTGGTGCTTCAAGCTGGTGCCCGTGGACCCCCGCGAGGTGAAGGAGGCCAACGAGGGC
GAGGACAACTGCCTGCTGCACCCCATGAGCCAGCACGGCGCCGAGGACGAGGACCGCGAGGTGCTGAAGTGGAAGTTCGA
CAGCCTGCTGGCCCACCGCCACATGGCCCGCGAGCTGCACCCCGAGTACTACAAGGACTGCTGA

FIGURE 26

>Nef_TV1_C_ZAwt (SEQ ID NO:56)

ATGGGAGGCAAGTGGTCAAAACGCAGCATAGTTGGATGGCCTGCAGTAAGAGAAAGAATGAGAAGAACTGAGCCAGCAGC
AGAGGGAGTAGGAGCAGCGTCTCAAGACTTAGATAGACATGGGGCACTTACAAGCAGCAACACACCTGCTACTAATGAAG
CTTGTGCCTGGCTGCAAGCACAAGAGGAGGACGGAGATGTAGGCTTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATG
ACTTATAAGAGTGCAGTAGATCTCAGCTTCTTTTTAAAAGAAAAGGGGGGACTGGAAGGGTTAATTTACTCTAGGAAAAG
GCAAGAAATCCTTGATTTGTGGGTCTATAACACACAAGGCTTCTTCCCTGATTGGCAAAACTACACATCGGGGCCAGGGG
TCCGATTCCCACTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGACCCAAGGGAGGTGAAAGAGGCCAATGAAGGA
GAAGACAACTGTTTGCTACACCCTATGAGCCAACATGGAGCAGAGGATGAAGATAGAGAAGTATTAAAGTGGAAGTTTGA
CAGCCTTCTAGCACACAGACACATGGCCCGCGAGCTACATCCGGAGTATTACAAAGACTGCTGA

FIGURE 27

>NefD125G_TV1_C_ZAopt (SEQ ID NO:57)

ATGGGCGGCAAGTGGAGCAAGCGCAGCATCGTGGGCTGGCCCGCCGTGCGCGAGCGCATGCGCCGCACCGAGCCCGCCGC
CGAGGGCGTGGGCGCCGCCAGCCAGGACCTGGACCGCCACGGCGCCCTGACCAGCAGCAACACCCCCGCCACCAACGAGG
CCTGCGCCTGGCTGCAGGCCCAGGAGGAGGACGGCGACGTGGGCTTCCCCGTGCGCCCCCAGGTGCCCCTGCGCCCCATG
ACCTACAAGAGCGCCGTGGACCTGAGCTTCTTCCTGAAGGAGAAGGGCGGCCTGGAGGGCCTGATCTACAGCCGCAAGCG
CCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGCTTCTTCCCCGGCTGGCAGAACTACACCAGCGGCCCCGGCG
TGCGCTTCCCCCTGACCTTCGGCTGGTGCTTCAAGCTGGTGCCCGTGGACCCCCGCGAGGTGAAGGAGGCCAACGAGGGC
GAGGACAACTGCCTGCTGCACCCCATGAGCCAGCACGGCGCCGAGGACGAGGACCGCGAGGTGCTGAAGTGGAAGTTCGA
CAGCCTGCTGGCCCACCGCCACATGGCCCGCGAGCTGCACCCCGAGTACTACAAGGACTGCTGA

FIGURE 28

>p15RNaseH_TV1_C_ZAopt (SEQ ID NO:58)

ACCTTCTACGTGGACGGCGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCTACGTGACCGACCGCGGCCGCCAGAA
GATCGTGACCCTGACCAACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCG
AGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGACAGCGAGATCTTC
AACCAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAA
CGAGCAGGTGGACAAGCTGGTGAGCAAGGGCATC

FIGURE 29

>p15RNaseH_TV1_C_Zawt (SEQ ID NO:59)

ACTTTCTATGTAGATGGAGCAACTAATAGGGAAGCTAAAATAGGAAAAGCAGGGTATGTTACTGACAGAGGAAGGCAGAA
AATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAAGCAATTCAGCTAGCTCTGCAGGATTCAGGATCAG
AAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAACCAGATAAGAGTGACTCAGAGATATTT
AACCAAATAATAGAACAGTTAATAAACAAGGAAAGAATCTACCTGTCATGGGTACCAGCACATAAAGGAATTGGGGGAAA
TGAACAAGTAGATAAATTAGTAAGTAAGGGAATT

FIGURE 30

>p31Int_TV1_C_Zaopt (SEQ ID NO:60)

CGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGCGCTACCACAGCAACTGGCGCGCCATGGC
CAACGAGTTCAACCTGCCCCCCATCGTGGCCAAGGAGATCGTGGCCAGCTGCGACAAGTGCCAGCTGAAGGGCGAGGCCA
TCCACGGCCAGGTGGACTGCAGCCCCGGCATCTGGCAGCTGGACTGCACCCACCTGGAGGGCAAGATCATCCTGGTGGCC
GTGCACGTGGCCAGCGGCTACATGGAGGCCGAGGTGATCCCCGCCGAGACCGGCCAGGAGACCGCCTACTTCATCCTGAA
GCTGGCCGGCCGCTGGCCCGTGAAGGTGATCCACACCGACAACGGCAGCAACTTCACCAGCACCGCCGTGAAGGCCGCCT
GCTGGTGGGCCGGCATCCAGCAGGAGTTCGGCATCCCCTACAACCCCCAGAGCCAGGGCGTGGTGGAGAGCATGAACAAG
GAGCTGAAGAAGATCATCGGCCAGGTGCGCGACCAGGCCGAGCACCTGAAGACCGCCGTGCAGATGGCCGTGTTCATCCA
CAACTTCAAGCGCAAGGGCGGCATCGGCGGCTACAGCGCCGGCGAGCGCATCATCGACATCATCGCCACCGACATCCAGA
CCAAGGAGCTGCAGAAGCAGATCATCCGCATCCAGAACTTCCGCGTGTACTACCGCGACAGCCGCGACCCCATCTGGAAG
GGCCCCGCCGAGCTGCTGTGGAAGGGCGAGGGCGTGGTGGTGATCGAGGACAAGGGCGACATCAAGGTGGTGCCCCGCCG
CAAGGCCAAGATCATCCGCGACTACGGCAAGCAGATGGCCGGCGCCGACTGCGTGGCCGGCGGCCAGGACGAGGAC

FIGURE 31

>p31Int_TV1_C_ZAwt (SEQ ID NO:61)

AGGAAAGTGTTGTTTCTAGATGGAATAGATAAAGCTCAAGAAGAGCATGAAAGGTACCACAGCAATTGGAGAGCAATGGC
TAATGAGTTTAATCTGCCACCCATAGTAGCAAAAGAAATAGTAGCTAGCTGTGATAAATGTCAGCTAAAAGGGGAAGCCA
TACATGGACAAGTCGACTGTAGTCCAGGGATATGGCAATTAGATTGTACCCATTTAGAGGGAAAAATCATCCTGGTAGCA
GTCCATGTAGCTAGTGGCTACATGGAAGCAGAGGTTATCCCAGCAGAAACAGGACAAGAAACAGCATATTTTATATTAAA
ATTAGCAGGAAGATGGCCAGTCAAAGTAATACATACAGACAATGGCAGTAATTTTACCAGTACTGCAGTTAAGGCAGCCT
GTTGGTGGGCAGGTATCCAACAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAGGGAGTGGTAGAATCCATGAATAAA
GAATTAAAGAAAATAATAGGACAAGTAAGAGATCAAGCTGAGCACCTTAAGACAGCAGTACAAATGGCAGTATTCATTCA
CAATTTTAAAAGAAAAGGGGGAATTGGGGGGTACAGTGCAGGGGAAAGAATAATAGACATAATAGCAACAGACATACAAA
CTAAAGAATTACAAAAACAAATTATAAGAATTCAAAATTTTCGGGTTTATTACAGAGACAGCAGAGACCCTATTTGGAAA
GGACCAGCCGAACTACTCTGGAAAGGTGAAGGGGTAGTAGTAATAGAAGATAAAGGTGACATAAAGGTAGTACCAAGGAG
GAAAGCAAAAATCATTAGAGATTATGGAAAACAGATGGCAGGTGCTGATTGTGTGGCAGGTGGACAGGATGAAGAT

FIGURE 32

>Pol_TV1_C_ZAopt (SEQ ID NO:62)

TTCTTCCGCGAGAACCTGGCCTTCCCCCAGGGCGAGGCCCGCGAGTTCCCCCCCGAGCAGACCCGCGCCAACAGCCCCAC
CAGCCGCACCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCGCCGAGGAGGGCGAGCGCGAGG
GCACCTTCAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCC
CTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATCGACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGG
CATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC
TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGC
CCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAA
GATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGACAACCCCTACA
ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGC
ACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGA
CGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGAGCTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA
ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGC
ATGACCAAGATCCTGGAGCCCTTCCGCGCCAAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGG
CAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGTGGGGCTTCACCACCC
CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCC
ATCCTGGTGCCCGAGAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGAT
CTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG
AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAG
GACCTGATCGCCGAGATCCAGAAGCAGGGCCACGAGCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAA
GACCGGCAAGTACGCCAAGATGCGCACCACCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA
TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACC
GACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGA
GAAGGACCCCATCGCCGGCGTGGAGACCTTCTACGTGGACGGCGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCT
ACGTGACCGACCGCGGCCGCCAGAAGATCGTGACCCTGACCAACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAG
CTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCC
CGACAAGAGCGACAGCGAGATCTTCAACCAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGC
CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGAC
GGCATCGACAAGGCCCAGGAGGAGCACGAGCGCTACCACAGCAACTGGCGCGCCATGGCCAACGAGTTCAACCTGCCCCC
CATCGTGGCCAAGGAGATCGTGGCCAGCTGCGACAAGTGCCAGCTGAAGGGCGAGGCCATCCACGGCCAGGTGGACTGCA
GCCCCGGCATCTGGCAGCTGGACTGCACCCACCTGGAGGGCAAGATCATCCTGGTGGCCGTGCACGTGGCCAGCGGCTAC
ATGGAGGCCGAGGTGATCCCCGCCGAGACCGGCCAGGAGACCGCCTACTTCATCCTGAAGCTGGCCGGCCGCTGGCCCGT
GAAGGTGATCCACACCGACAACGGCAGCAACTTCACCAGCACCGCCGTGAAGGCCGCCTGCTGGTGGGCCGGCATCCAGC
AGGAGTTCGGCATCCCCTACAACCCCCAGAGCCAGGGCGTGGTGGAGAGCATGAACAAGGAGCTGAAGAAGATCATCGGC
CAGGTGCGCGACCAGGCCGAGCACCTGAAGACCGCCGTGCAGATGGCCGTGTTCATCCACAACTTCAAGCGCAAGGGCGG
CATCGGCGGCTACAGCGCCGGCGAGCGCATCATCGACATCATCGCCACCGACATCCAGACCAAGGAGCTGCAGAAGCAGA
TCATCCGCATCCAGAACTTCCGCGTGTACTACCGCGACAGCCGCGACCCCATCTGGAAGGGCCCCGCCGAGCTGCTGTGG
AAGGGCGAGGGCGTGGTGGTGATCGAGGACAAGGGCGACATCAAGGTGGTGCCCCGCCGCAAGGCCAAGATCATCCGCGA
CTACGGCAAGCAGATGGCCGGCGCCGACTGCGTGGCCGGCGGCCAGGACGAGGAC

FIGURE 33

>Pol_TV1_C_ZAwt (SEQ ID NO:63)

TTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGGAGGCCAGGGAATTTCCTCCAGAACAGACCAGAGCCAACAGCCCCAC
CAGCAGAACCAACAGCCCCACCAGCAGAGAGCTTCAGGTTCGAGGAGACAACCCCCGTGCCGAGGAAGGAGAAAGAGAGG
GAACCTTTAACTTCCCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCT
CTCTTAGACACAGGAGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGGAAACCAAAAATGATAGGGGG
AATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCTATAGGTACAGTAT
TAGTAGGGCCTACACCAGTCAACATAATTGGAAGAAATCTGTTAACTCAGCTTGGATGCACACTAAATTTTCCAATTAGT
CCTATTGAAACTGTACCAGTAAAATTAAAACCAGGAATGGATGGCCCAAAGGTCAAACAATGGCCATTGACAGAAGAAAA
AATAAAAGCATTAACAGCAATTTGTGAGGAAATGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGATAATCCATATA
ACACTCCAGTATTTGCCATAAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAGGGAACTCAATAAAAGA
ACTCAAGACTTTTGGGAAGTTCAATTAGGAATACCACACCCAGCAGGATTAAAAAAGAAAAAATCAGTGACAGTGCTAGA
TGTGGGGGATGCATATTTTTCAGTTCCTTTAGATGAAAGCTTCAGGAAATATACTGCATTCACCATACCTAGTATAAACA
ATGAAACACCAGGGATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAGAGTAGC
ATGACAAAAATCTTAGAGCCCTTCAGAGCAAAAAATCCAGACATAGTTATCTATCAATATATGGATGACTTGTATGTAGG
ATCTGACTTAGAAATAGGGCAACATAGAGCAAAAATAGAAGAGTTAAGGGAACATTTATTGAAATGGGGATTTACAACAC
CAGACAAGAAACATCAAAAAGAACCCCCATTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACAGTACAACCT
ATACTGCTGCCAGAAAAGGATAGTTGGACTGTCAATGATATACAGAAGTTAGTGGGAAAATTAAACTGGGCAAGTCAGAT
TTACCCAGGGATTAAAGTAAGGCAACTCTGTAAACTCCTCAGGGGGGCCAAAGCACTAACAGACATAGTACCACTAACTG
AAGAAGCAGAATTAGAATTGGCAGAGAACAGGGAAATTTTAAGAGAACCAGTACATGGAGTATATTATGATCCATCAAAA
GACTTGATAGCTGAAATACAGAAACAGGGGCATGAACAATGGACATATCAAATTTATCAAGAACCATTTAAAAATCTGAA
AACAGGGAAGTATGCAAAAATGAGGACTACCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCCA
TGGAAAGCATAGTAATATGGGGAAAGACTCCTAAATTTAGACTACCCATCCAAAAAGAAACATGGGAGACATGGTGGACA
GACTATTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTTGTTAATACCCCTCCCCTAGTAAAATTATGGTACCAACTAGA
AAAAGATCCCATAGCAGGAGTAGAAACTTTCTATGTAGATGGAGCAACTAATAGGGAAGCTAAAATAGGAAAAGCAGGGT
ATGTTACTGACAGAGGAAGGCAGAAAATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAAGCAATTCAG
CTAGCTCTGCAGGATTCAGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAACC
AGATAAGAGTGACTCAGAGATATTTAACCAAATAATAGAACAGTTAATAAACAAGGAAAGAATCTACCTGTCATGGGTAC
CAGCACATAAAGGAATTGGGGGAAATGAACAAGTAGATAAATTAGTAAGTAAGGGAATTAGGAAAGTGTTGTTTCTAGAT
GGAATAGATAAAGCTCAAGAAGAGCATGAAAGGTACCACAGCAATTGGAGAGCAATGGCTAATGAGTTTAATCTGCCACC
CATAGTAGCAAAAGAAATAGTAGCTAGCTGTGATAAATGTCAGCTAAAAGGGGAAGCCATACATGGACAAGTCGACTGTA
GTCCAGGGATATGGCAATTAGATTGTACCCATTTAGAGGGAAAAATCATCCTGGTAGCAGTCCATGTAGCTAGTGGCTAC
ATGGAAGCAGAGGTTATCCCAGCAGAAACAGGACAAGAAACAGCATATTTTATATTAAAATTAGCAGGAAGATGGCCAGT
CAAAGTAATACATACAGACAATGGCAGTAATTTTACCAGTACTGCAGTTAAGGCAGCCTGTTGGTGGGCAGGTATCCAAC
AGGAATTTGGAATTCCCTACAATCCCCAAAGTCAGGGAGTGGTAGAATCCATGAATAAAGAATTAAAGAAAATAATAGGA
CAAGTAAGAGATCAAGCTGAGCACCTTAAGACAGCAGTACAAATGGCAGTATTCATTCACAATTTTAAAAGAAAAGGGGG
AATTGGGGGGTACAGTGCAGGGGAAAGAATAATAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAA
TTATAAGAATTCAAAATTTTCGGGTTTATTACAGAGACAGCAGAGACCCTATTTGGAAAGGACCAGCCGAACTACTCTGG
AAAGGTGAAGGGGTAGTAGTAATAGAAGATAAAGGTGACATAAAGGTAGTACCAAGGAGGAAAGCAAAAATCATTAGAGA
TTATGGAAAACAGATGGCAGGTGCTGATTGTGTGGCAGGTGGACAGGATGAAGAT

FIGURE 34

>Prot_TV1_C_ZAopt (SEQ ID NO:64)

CCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCCCTGCTGGACACCGG
CGCCGACGACACCGTGCTGGAGGAGATCGACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA
TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCTGAACTTC

FIGURE 35

>Prot_TV1_C_ZAwt (SEQ ID NO:65)

CCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCTCTCTTAGACACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTA
TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGGCCTACA
CCAGTCAACATAATTGGAAGAAATCTGTTAACTCAGCTTGGATGCACACTAAATTTT

FIGURE 36

>Protina_TV1_C_ZAopt (SEQ ID NO:66)

CCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCCCTGCTGGCCACCGG
CGCCGACGACACCGTGCTGGAGGAGATCGACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA
TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCTGAACTTC

FIGURE 37

>Protina_TV1_C_ZAwt (SEQ ID NO:67)

CCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCTCTCTTAGCCACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTA
TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGGCCTACA
CCAGTCAACATAATTGGAAGAAATCTGTTAACTCAGCTTGGATGCACACTAAATTTT

FIGURE 38

>ProtinaRTmut_TV1_C_ZAopt (SEQ ID NO:68)

CCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCCCTGCTGGCCACCGG
CGCCGACGACACCGTGCTGGAGGAGATCGACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA
TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT
GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA
CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGACAACCCCTACAACACCCCCGTGTTC
GCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG
GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT
ACTTCAGCGTGCCCCTGGACGAGAGCTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGC
ATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCT
GGAGCCCTTCCGCGCCAAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC
AGCACCGCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAG
GAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCCTGCTGCCCGAGAAGGACAGCTG
GACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGC
TGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAG
AACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCA
GGGCCACGAGCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCA
CCACCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG
ACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCC
CGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGACCCCATCGCCGGCGTGGAGA
CCTTCTACGTGGACGGCGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCTACGTGACCGACCGCGGCCGCCAGAAG
ATCGTGACCCTGACCAACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGA
GGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGACAGCGAGATCTTCA
ACCAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAAC
GAGCAGGTGGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG

FIGURE 39

>ProtinaRTmut_TV1_C_ZAwt (SEQ ID NO:69)

CCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCTCTCTTAGCCACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTA
TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGGCCTACA
CCAGTCAACATAATTGGAAGAAATCTGTTAACTCAGCTTGGATGCACACTAAATTTTCCAATTAGTCCTATTGAAACTGT
ACCAGTAAAATTAAAACCAGGAATGGATGGCCCAAAGGTCAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAA
CAGCAATTTGTGAGGAAATGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGATAATCCATATAACACTCCAGTATTT
GCCATAAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAGGGAACTCAATAAAAGAACTCAAGACTTTTG
GGAAGTTCAATTAGGAATACCACACCCAGCAGGATTAAAAAAGAAAAAATCAGTGACAGTGCTAGATGTGGGGGATGCAT
ATTTTTCAGTTCCTTTAGATGAAAGCTTCAGGAAATATACTGCATTCACCATACCTAGTATAAACAATGAAACACCAGGG
ATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAGAGTAGCATGACAAAAATCTT
AGAGCCCTTCAGAGCAAAAAATCCAGACATAGTTATCTATCAAGCCCCGTTGTATGTAGGATCTGACTTAGAAATAGGGC
AACATAGAGCAAAAATAGAAGAGTTAAGGGAACATTTATTGAAATGGGGATTTACAACACCAGACAAGAAACATCAAAAA
GAACCCCCATTTCTTCCCATCGAACTCCATCCTGACAAATGGACAGTACAACCTATACTGCTGCCAGAAAAGGATAGTTG
GACTGTCAATGATATACAGAAGTTAGTGGGAAAATTAAACTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAAC
TCTGTAAACTCCTCAGGGGGGCCAAAGCACTAACAGACATAGTACCACTAACTGAAGAAGCAGAATTAGAATTGGCAGAG
AACAGGGAAATTTTAAGAGAACCAGTACATGGAGTATATTATGATCCATCAAAAGACTTGATAGCTGAAATACAGAAACA
GGGGCATGAACAATGGACATATCAAATTTATCAAGAACCATTTAAAAATCTGAAAACAGGGAAGTATGCAAAAATGAGGA
CTACCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCCATGGAAAGCATAGTAATATGGGGAAAG
ACTCCTAAATTTAGACTACCCATCCAAAAAGAAACATGGGAGACATGGTGGACAGACTATTGGCAAGCCACCTGGATCCC
TGAGTGGGAGTTTGTTAATACCCCTCCCCTAGTAAAATTATGGTACCAACTAGAAAAAGATCCCATAGCAGGAGTAGAAA
CTTTCTATGTAGATGGAGCAACTAATAGGGAAGCTAAAATAGGAAAAGCAGGGTATGTTACTGACAGAGGAAGGCAGAAA
ATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAAGCAATTCAGCTAGCTCTGCAGGATTCAGGATCAGA
AGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAACCAGATAAGAGTGACTCAGAGATATTTA
ACCAAATAATAGAACAGTTAATAAACAAGGAAAGAATCTACCTGTCATGGGTACCAGCACATAAAGGAATTGGGGGAAAT
GAACAAGTAGATAAATTAGTAAGTAAGGGAATTAGGAAAGTGTTG

FIGURE 40

>ProtwtRTwt_TV1_C_ZAopt (SEQ ID NO:70)

CCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGAGGGCCAGATCAAGGAGGCCCTGCTGGACACCGG
CGCCGACGACACCGTGCTGGAGGAGATCGACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA
TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGGTGGGCCCCACC
CCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT
GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA
CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGACAACCCCTACAACACCCCCGTGTTC
GCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG
GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT
ACTTCAGCGTGCCCCTGGACGAGAGCTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGC
ATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCT
GGAGCCCTTCCGCGCCAAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGA
TCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGTGGGGCTTCACCACCCCCGACAAGAAGCAC
CAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCCTGCTGCCCGA
GAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCA
AGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTG
GAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGA
GATCCAGAAGCAGGGCCACGAGCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACG
CCAAGATGCGCACCACCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTG
ATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGC
CACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGACCCCATCG
CCGGCGTGGAGACCTTCTACGTGGACGGCGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCTACGTGACCGACCGC
GGCCGCCAGAAGATCGTGACCCTGACCAACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGA
CAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGACA
GCGAGATCTTCAACCAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCCGCCCACAAGGGC
ATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG

FIGURE 41

>ProtwtRTwt_TV1_C_ZAwt(SEQ ID NO:71)

CCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAAAAGTAGAGGGCCAGATAAAGGAGGCTCTCTTAGACACAGG
AGCAGATGATACAGTATTAGAAGAAATAGATTTGCCAGGGAAATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTA
TCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCTATAGGTACAGTATTAGTAGGGCCTACA
CCAGTCAACATAATTGGAAGAAATCTGTTAACTCAGCTTGGATGCACACTAAATTTTCCAATTAGTCCTATTGAAACTGT
ACCAGTAAAATTAAAACCAGGAATGGATGGCCCAAAGGTCAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAA
CAGCAATTTGTGAGGAAATGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGATAATCCATATAACACTCCAGTATTT
GCCATAAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAGGGAACTCAATAAAAGAACTCAAGACTTTTG
GGAAGTTCAATTAGGAATACCACACCCAGCAGGATTAAAAAAGAAAAAATCAGTGACAGTGCTAGATGTGGGGGATGCAT
ATTTTTCAGTTCCTTTAGATGAAAGCTTCAGGAAATATACTGCATTCACCATACCTAGTATAAACAATGAAACACCAGGG
ATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAGAGTAGCATGACAAAAATCTT
AGAGCCCTTCAGAGCAAAAAATCCAGACATAGTTATCTATCAATATATGGATGACTTGTATGTAGGATCTGACTTAGAAA
TAGGGCAACATAGAGCAAAAATAGAAGAGTTAAGGGAACATTTATTGAAATGGGGATTTACAACACCAGACAAGAAACAT
CAAAAAGAACCCCCATTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACAGTACAACCTATACTGCTGCCAGA
AAAGGATAGTTGGACTGTCAATGATATACAGAAGTTAGTGGGAAAATTAAACTGGGCAAGTCAGATTTACCCAGGGATTA
AAGTAAGGCAACTCTGTAAACTCCTCAGGGGGGCCAAAGCACTAACAGACATAGTACCACTAACTGAAGAAGCAGAATTA
GAATTGGCAGAGAACAGGGAAATTTTAAGAGAACCAGTACATGGAGTATATTATGATCCATCAAAAGACTTGATAGCTGA
AATACAGAAACAGGGGCATGAACAATGGACATATCAAATTTATCAAGAACCATTTAAAAATCTGAAAACAGGGAAGTATG
CAAAAATGAGGACTACCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCCATGGAAAGCATAGTA
ATATGGGGAAAGACTCCTAAATTTAGACTACCCATCCAAAAAGAAACATGGGAGACATGGTGGACAGACTATTGGCAAGC
CACCTGGATCCCTGAGTGGGAGTTTGTTAATACCCCTCCCCTAGTAAAATTATGGTACCAACTAGAAAAAGATCCCATAG
CAGGAGTAGAAACTTTCTATGTAGATGGAGCAACTAATAGGGAAGCTAAAATAGGAAAAGCAGGGTATGTTACTGACAGA
GGAAGGCAGAAAATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAAGCAATTCAGCTAGCTCTGCAGGA
TTCAGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAACCAGATAAGAGTGACT
CAGAGATATTTAACCAAATAATAGAACAGTTAATAAACAAGGAAAGAATCTACCTGTCATGGGTACCAGCACATAAAGGA
ATTGGGGGAAATGAACAAGTAGATAAATTAGTAAGTAAGGGAATTAGGAAAGTGTTG

FIGURE 42

>RevExon1_TV1_C_ZAopt(SEQ ID NO:72)

ATGGCCGGCCGCAGCGGCGACAGCGACGAGGCCCTGCTGCAGGTGGTGAAGATCATCAAGATCCTGTACCAGAGC

FIGURE 43

>RevExon1_TV1_C_ZAwt(SEQ ID NO:73)

ATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGCTCCTCCAAGTGGTGAAGATCATCAAAATCCTCTATCAAAGCA

FIGURE 44

>RevExon2_TV1_C_ZAopt-2(SEQ ID NO:74)

CCCTACCCCAAGCCCGAGGGCACCCGCCAGGCCCGCCGCAACCGCCGCCGCCGCTGGCGCGCCCGCCAGCGCCAGATCCA
CACCATCGGCGAGCGCATCCTGGTGGCCTGCCTGGGCCGCAGCGCCGAGCCCGTGCCCCTGCAGCTGCCCCCCCTGGAGC
GCCTGCACATCAACTGCAGCGAGGGCAGCGGCACCAGCGGCACCCAGCAGAGCCAGGGCACCACCGAGGGCGTGGGCGAC
CCCTAA

FIGURE 45

>RevExon2_TV1_C_ZAwt(SEQ ID NO:75)

ACCCTTACCCCAAGCCCGAGGGGACTCGACAGGCTCGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGACAGAGACAGATC
CATACGATTGGTGAGCGGATTCTTGTCGCTTGCCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGA
GAGACTTCATATTAATTGCAGTGAGGGCAGTGGAACTTCTGGGACACAGCAGTCTCAGGGGACTACAGAGGGGGTGGGAG
ATCCTTAA

FIGURE 46

RT_TV1_C_ZAopt (SEQ ID NO:76)

CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCA
AGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCG
AGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGACAACCCCTACAACA
CCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTT
CCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC
CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC
TTCAGCGTGCCCCTGGACGAGAGCTTCCGCAAGTACACCGCCTTCACCATCCCCAGCA
TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAA
GGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC
AAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACC
TGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGT
GGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGG
CTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCCTGCTGCCCGAGAAGGAC
AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAG
ATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC
TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCG
AGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGATCGC
CGAGATCCAGAAGCAGGGCCACGAGCAGTGGACCTACCAGATCTACCAGGAGCCCTT
CAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCACCCACACCAACGACGT
GAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGG
CAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACC
GACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG
TGAAGCTGTGGTACCAGCTGGAGAAGGACCCCATCGCCGGCGTGGAGACCTTCTACGT
GGACGGCGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCTACGTGACCGACCG
CGGCCGCCAGAAGATCGTGACCCTGACCAACACCACCAACCAGAAGACCGAGCTGCA
GGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAG
CCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGACAGCGAGATCTTC
AACCAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCCG
CCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGAGCAAGGGCATCC
GCAAGGTGCTG

FIGURE 47

>RT_TV1_C_ZAwt(SEQ ID NO:77)

CCAATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAACCAGGAATGGATGGCCCAAAGGTCAAACAATGGCCATTGAC
AGAAGAAAAAATAAAAGCATTAACAGCAATTTGTGAGGAAATGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGATA
ATCCATATAACACTCCAGTATTTGCCATAAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAGGGAACTC
AATAAAAGAACTCAAGACTTTTGGGAAGTTCAATTAGGAATACCACACCCAGCAGGATTAAAAAAGAAAAAATCAGTGAC
AGTGCTAGATGTGGGGGATGCATATTTTTCAGTTCCTTTAGATGAAAGCTTCAGGAAATATACTGCATTCACCATACCTA
GTATAAACAATGAAACACCAGGGATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTC
CAGAGTAGCATGACAAAAATCTTAGAGCCCTTCAGAGCAAAAAATCCAGACATAGTTATCTATCAATATATGGATGACTT
GTATGTAGGATCTGACTTAGAAATAGGGCAACATAGAGCAAAAATAGAAGAGTTAAGGGAACATTTATTGAAATGGGGAT
TTACAACACCAGACAAGAAACATCAAAAAGAACCCCCATTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACA
GTACAACCTATACTGCTGCCAGAAAAGGATAGTTGGACTGTCAATGATATACAGAAGTTAGTGGGAAAATTAAACTGGGC
AAGTCAGATTTACCCAGGGATTAAAGTAAGGCAACTCTGTAAACTCCTCAGGGGGGCCAAAGCACTAACAGACATAGTAC
CACTAACTGAAGAAGCAGAATTAGAATTGGCAGAGAACAGGGAAATTTTAAGAGAACCAGTACATGGAGTATATTATGAT
CCATCAAAAGACTTGATAGCTGAAATACAGAAACAGGGGCATGAACAATGGACATATCAAATTTATCAAGAACCATTTAA
AAATCTGAAAACAGGGAAGTATGCAAAAATGAGGACTACCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAA
AAATAGCCATGGAAAGCATAGTAATATGGGGAAAGACTCCTAAATTTAGACTACCCATCCAAAAAGAAACATGGGAGACA
TGGTGGACAGACTATTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTTGTTAATACCCCTCCCCTAGTAAAATTATGGTA
CCAACTAGAAAAAGATCCCATAGCAGGAGTAGAAACTTTCTATGTAGATGGAGCAACTAATAGGGAAGCTAAAATAGGAA
AAGCAGGGTATGTTACTGACAGAGGAAGGCAGAAAATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAA
GCAATTCAGCTAGCTCTGCAGGATTCAGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCA
AGCACAACCAGATAAGAGTGACTCAGAGATATTTAACCAAATAATAGAACAGTTAATAAACAAGGAAAGAATCTACCTGT
CATGGGTACCAGCACATAAAGGAATTGGGGGAAATGAACAAGTAGATAAATTAGTAAGTAAGGGAATTAGGAAAGTGTTG

FIGURE 48

>RTmut_TV1_C_ZAopt(SEQ ID NO:78)

CCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGAC
CGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGACA
ACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG
AACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGAC
CGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGAGCTTCCGCAAGTACACCGCCTTCACCATCCCCA
GCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTC
CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCAAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGT
GGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCGAGCACCTGCTGAAGTGGGGCTTCACCA
CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATC
CTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTA
CCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGG
AGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGAC
CTGATCGCCGAGATCCAGAAGCAGGGCCACGAGCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGAC
CGGCAAGTACGCCAAGATGCGCACCACCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGG
AGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGAC
TACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAA
GGACCCCATCGCCGGCGTGGAGACCTTCTACGTGGACGGCGCCACCAACCGCGAGGCCAAGATCGGCAAGGCCGGCTACG
TGACCGACCGCGGCCGCCAGAAGATCGTGACCCTGACCAACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTG
GCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGA
CAAGAGCGACAGCGAGATCTTCAACCAGATCATCGAGCAGCTGATCAACAAGGAGCGCATCTACCTGAGCTGGGTGCCCG
CCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG

FIGURE 49

>RTmut_TV1_C_ZAwt(SEQ ID NO:79)

CCAATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAACCAGGAATGGATGGCCCAAAGGTCAAACAATGGCCATTGAC
AGAAGAAAAAATAAAAGCATTAACAGCAATTTGTGAGGAAATGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGATA
ATCCATATAACACTCCAGTATTTGCCATAAAAAAGAAGGACAGTACTAAGTGGAGAAAATTAGTAGATTTCAGGGAACTC
AATAAAAGAACTCAAGACTTTTGGGAAGTTCAATTAGGAATACCACACCCAGCAGGATTAAAAAAGAAAAAATCAGTGAC
AGTGCTAGATGTGGGGGATGCATATTTTTCAGTTCCTTTAGATGAAAGCTTCAGGAAATATACTGCATTCACCATACCTA
GTATAAACAATGAAACACCAGGGATTAGATATCAATATAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTC
CAGAGTAGCATGACAAAAATCTTAGAGCCCTTCAGAGCAAAAAATCCAGACATAGTTATCTATCAAGCCCCGTTGTATGT
AGGATCTGACTTAGAAATAGGGCAACATAGAGCAAAAATAGAAGAGTTAAGGGAACATTTATTGAAATGGGGATTTACAA
CACCAGACAAGAAACATCAAAAAGAACCCCCATTTCTTCCCATCGAACTCCATCCTGACAAATGGACAGTACAACCTATA
CTGCTGCCAGAAAAGGATAGTTGGACTGTCAATGATATACAGAAGTTAGTGGGAAAATTAAACTGGGCAAGTCAGATTTA
CCCAGGGATTAAAGTAAGGCAACTCTGTAAACTCCTCAGGGGGGCCAAAGCACTAACAGACATAGTACCACTAACTGAAG
AAGCAGAATTAGAATTGGCAGAGAACAGGGAAATTTTAAGAGAACCAGTACATGGAGTATATTATGATCCATCAAAAGAC
TTGATAGCTGAAATACAGAAACAGGGGCATGAACAATGGACATATCAAATTTATCAAGAACCATTTAAAAATCTGAAAAC
AGGGAAGTATGCAAAAATGAGGACTACCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAATAGCCATGG
AAAGCATAGTAATATGGGGAAAGACTCCTAAATTTAGACTACCCATCCAAAAAGAAACATGGGAGACATGGTGGACAGAC
TATTGGCAAGCCACCTGGATCCCTGAGTGGGAGTTTGTTAATACCCCTCCCCTAGTAAAATTATGGTACCAACTAGAAAA
AGATCCCATAGCAGGAGTAGAAACTTTCTATGTAGATGGAGCAACTAATAGGGAAGCTAAAATAGGAAAAGCAGGGTATG
TTACTGACAGAGGAAGGCAGAAAATTGTTACTCTAACTAACACAACAAATCAGAAGACTGAGTTACAAGCAATTCAGCTA
GCTCTGCAGGATTCAGGATCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAACCAGA
TAAGAGTGACTCAGAGATATTTAACCAAATAATAGAACAGTTAATAAACAAGGAAAGAATCTACCTGTCATGGGTACCAG
CACATAAAGGAATTGGGGGAAATGAACAAGTAGATAAATTAGTAAGTAAGGGAATTAGGAAAGTGTTG

FIGURE 50

>TatC22Exon1_TV1_C_ZAopt(SEQ ID NO:80)

ATGGAGCCCGTGGACCCCAAGCTGAAGCCCTGGAACCACCCCGGCAGCCAGCCCAAGACCGCCGGCAACAACTGCTTCTG
CAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCCAGACCAAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCC
AGCGCCGCAGCGCCCCCCCCAGCGGCGAGGACCACCAGAACCCCCTGAGCAAGCAG

FIGURE 51

>TatExon1_TV1_C_ZAopt(SEQ ID NO:81)

ATGGAGCCCGTGGACCCCAAGCTGAAGCCCTGGAACCACCCCGGCAGCCAGCCCAAGACCGCCTGCAACAACTGCTTCTG
CAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCCAGACCAAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCC
AGCGCCGCAGCGCCCCCCCCAGCGGCGAGGACCACCAGAACCCCCTGAGCAAGCAG

FIGURE 52

>TatExon1_TV1_C_ZAwt(SEQ ID NO:82)

ATGGAGCCAGTAGATCCTAAACTAAAGCCCTGGAACCATCCAGGAAGCCAACCTAAAACAGCTTGTAATAATTGCTTTTG
CAAACACTGTAGCTATCATTGTCTAGTTTGCTTTCAGACAAAAGGTTTAGGCATTTCCTATGGCAGGAAGAAGCGGAGAC
AGCGACGAAGCGCTCCTCCAAGTGGTGAAGATCATCAAAATCCTCTATCAAAGCAG

FIGURE 53

>TatExon2_TV1_C_ZAopt(SEQ ID NO:83)

CCCCTGCCCCAGGCCCGCGGCGACAGCACCGGCAGCGAGGAGAGCAAGAAGAAGGTGGAGAGCAAGACCGAGACCGACCC
CTACGACTGGTGA

FIGURE 54

>TatExon2_TV1_C_ZAwt(SEQ ID NO:84)

CCCTTACCCCAAGCCCGAGGGGACTCGACAGGCTCGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGACAGAGACAGATCC
ATACGATTGGTGA

FIGURE 55

>Vif_TV1_C_ZAopt(SEQ ID NO:85)

ATGGAGAACCGCTGGCAGGTGCTGATCGTGTGGCAGGTGGACCGCATGAAGATCCGCGCCTGGAACAGCCTGGTGAAGCA
CCACATGTACATCAGCCGCCGCGCCAGCGGCTGGGTGTACCGCCACCACTTCGAGAGCCGCCACCCCAAGGTGAGCAGCG
AGGTGCACATCCCCCTGGGCGACGCCCGCCTGGTGATCAAGACCTACTGGGGCCTGCAGACCGGCGAGCGCGACTGGCAC
CTGGGCCACGGCGTGAGCATCGAGTGGCGCCTGCGCGAGTACAGCACCCAGGTGGACCCCGACCTGGCCGACCAGCTGAT
CCACATGCACTACTTCGACTGCTTCACCGAGAGCGCCATCCGCCAGGCCATCCTGGGCCACATCGTGTTCCCCCGCTGCG
ACTACCAGGCCGGCCACAAGAAGGTGGGCAGCCTGCAGTACCTGGCCCTGACCGCCCTGATCAAGCCCAAGAAGCGCAAG
CCCCCCCTGCCCAGCGTGCGCAAGCTGGTGGAGGACCGCTGGAACGACCCCCAGAAGACCCGCGGCCGCCGCGGCAACCA
CACCATGAACGGCCACTAG

FIGURE 56

f_TV1_C_ZAwt(SEQ ID NO:86)

GAAAACAGATGGCAGGTGCTGATTGTGTGGCAGGTGGACAGGATGAAGATTAGAGCATGGAATAGTTTAGTAAAGCA
TATGTATATATCAAGGAGAGCTAGTGGATGGGTCTACAGACATCATTTTGAAAGCAGACATCCAAAAGTAAGTTCAG
TACATATCCCATTAGGGGATGCTAGATTAGTAATAAAAACATATTGGGGTTTGCAGACAGGAGAAAGAGATTGGCAT
GGTCATGGAGTCTCCATAGAATGGAGACTGAGAGAATACAGCACACAAGTAGACCCTGACCTGGCAGACCAGCTAAT
CATGCATTATTTTGATTGTTTTACAGAATCTGCCATAAGACAAGCCATATTAGGACACATAGTTTTTCCTAGGTGTG
ATCAAGCAGGACATAAGAAGGTAGGATCTCTGCAATACTTGGCACTGACAGCATTGATAAAACCAAAAAAGAGAAAG
CCTCTGCCTAGTGTTAGAAAATTAGTAGAGGATAGATGGAACGACCCCCAGAAGACCAGGGGCCGCAGAGGGAACCA
AATGAATGGACACTAG

FIGURE 57

>Vpr_TV1_C_ZAopt(SEQ ID NO:87)

ATGGAGCGCCCCCCCGAGGACCAGGGCCCCCAGCGCGAGCCCTACAACGAGTGGACCCTGGAGATCCTGGAGGAGCTGAA
GCAGGAGGCCGTGCGCCACTTCCCCCGCCCCTGGCTGCACAGCCTGGGCCAGTACATCTACGAGACCTACGGCGACACCT
GGACCGGCGTGGAGGCCATCATCCGCGTGCTGCAGCAGCTGCTGTTCATCCACTTCCGCATCGGCTGCCAGCACAGCCGC
ATCGGCATCCTGCGCCAGCGCCGCGCCCGCAACGGCGCCAGCCGCAGC

FIGURE 58

>Vpr_TV1_C_ZAwt(SEQ ID NO:88)

ATGGAACGACCCCCAGAAGACCAGGGGCCGCAGAGGGAACCATACAATGAATGGACACTAGAGATTCTAGAAGAACTCAA
GCAGGAAGCTGTCAGACACTTTCCTAGACCATGGCTCCATAGCTTAGGACAATATATCTATGAAACCTATGGGGATACTT
GGACGGGAGTTGAAGCTATAATAAGAGTACTGCAACAACTACTGTTCATTCATTTCAGAATTGGATGCCAACATAGCAGA
ATAGGCATCTTGCGACAGAGAAGAGCAAGAAATGGAGCCAGTAGATCC

FIGURE 59

>Vpu_TV1_C_ZAopt(SEQ ID NO:89)

ATGGTGAGCCTGAGCCTGTTCAAGGGCGTGGACTACCGCCTGGGCGTGGGCGCCCTGATCGTGGCCCTGATCATCGCCAT
CATCGTGTGGACCATCGCCTACATCGAGTACCGCAAGCTGGTGCGCCAGAAGAAGATCGACTGGCTGATCAAGCGCATCC
GCGAGCGCGCCGAGGACAGCGGCAACGAGAGCGACGGCGACACCGAGGAGCTGAGCACCATGGTGGACATGGGCCACCTG
CGCCTGCTGGACGCCAACGACCTGTAA

FIGURE 60

>Vpu_TV1_C_ZAwt(SEQ ID NO:90)

ATGGTAAGTTTAAGTTTATTTAAAGGAGTAGATTATAGATTAGGAGTAGGAGCATTGATAGTAGCACTAATCATAGCAAT
AATAGTGTGGACCATAGCATATATAGAATATAGGAAATTGGTAAGACAAAAGAAAATAGACTGGTTAATTAAAAGAATTA
GGGAAAGAGCAGAAGACAGTGGCAATGAGAGTGATGGGGACACAGAAGAATTGTCAACAATGGTGGATATGGGGCATCTT
AGGCTTCTGGATGCTAATGATTTGTAA

FIGURE 61

dna revexon1_2TV1_C_ZAop (SEQ ID NO:91)

ATGGCCGGCCGCAGCGGCGACAGCGACGAGGCCCTGCTGCAGGTGGTGAAGATCATC
AAGATCCTGTACCAGAGCCCCTACCCCAAGCCCGAGGGCACCCGCCAGGCCCGCCGCA
ACCGCCGCCGCCGCTGGCGCGCCCGCCAGCGCCAGATCCACACCATCGGCGAGCGCAT
CCTGGTGGCCTGCCTGGGCCGCAGCGCCGAGCCCGTGCCCCTGCAGCTGCCCCCCCTG
GAGCGCCTGCACATCAACTGCAGCGAGGGCAGCGGCACCAGCGGCACCCAGCAGAGC
CAGGGCACCACCGAGGGCGTGGGCGACCCCTAA

FIGURE 62

dna Revexon1_2_TV1_C_ZAwt (SEQ ID NO:92)

ATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGCTCCTCCAAGTGGTGAAGATCATC
AAAATCCTCTATCAAAGCAACCCTTACCCCAAGCCCGAGGGGACTCGACAGGCTCGGA
GGAATCGAAGAAGAAGGTGGAGAGCAAGACAGAGACAGATCCATACGATTGGTGAGC
GGATTCTTGTCGCTTGCCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCG
CTTGAGAGACTTCATATTAATTGCAGTGAGGGCAGTGGAACTTCTGGGACACAGCAGT
CTCAGGGGACTACAGAGGGGGTGGGAGATCCTTAA

FIGURE 63

dna TatC22Exon1_2_TV1_C_ZAopt (SEQ ID NO:93)

ATGGAGCCCGTGGACCCCAAGCTGAAGCCCTGGAACCACCCCGGCAGCCAGCCCAAG
ACCGCCGGCAACAACTGCTTCTGCAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCC
AGACCAAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCCAGCGCCGCAGCG
CCCCCCCCCAGCGGCGAGGACCACCAGAACCCCCTGAGCAAGCAGCCCCTGCCCCAGGC
CCGCGGCGACAGCACCGGCAGCGAGGAGAGCAAGAAGAAGGTGGAGAGCAAGACCG
AGACCGACCCCTACGACTGGTGA

FIGURE 64

dna TatExon1_2_TV1_C_ZAopt (SEQ ID NO:94)

ATGGAGCCCGTGGACCCCAAGCTGAAGCCCTGGAACCACCCCGGCAGCCAGCCCAAG
ACCGCCTGCAACAACTGCTTCTGCAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCCA
GACCAAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCAGCGCCGCAGCGCC
CCCCCCAGCGGCGAGGACCACCAGAACCCCCTGAGCAAGCAGCCCCTGCCCCAGGCCC
GCGGCGACAGCACCGGCAGCGAGGAGAGCAAGAAGAAGGTGGAGAGCAAGACCGAG
ACCGACCCCTACGACTGGTGA

FIGURE 65

dna TatExon1_2_TV1_C_ZAwt (SEQ ID NO:95)

ATGGAGCCAGTAGATCCTAAACTAAAGCCCTGGAACCATCCAGGAAGCCAACCTAAA
ACAGCTTGTAATAATTGCTTTTGCAAACACTGTAGCTATCATTGTCTAGTTTGCTTTCA
GACAAAAGGTTTAGGCATTTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGC
TCCTCCAAGTGGTGAAGATCATCAAAATCCTCTATCAAAGCAGCCCTTACCCCAAGCC
CGAGGGGACTCGACAGGCTCGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGACAGA
GACAGATCCATACGATTGGTGA

FIGURE 66

NefD125G-Myr_TV1_C_ZAopt (SEQ ID NO:96)

ATGGCCGGCAAGTGGAGCAAGCGCAGCATCGTGGGCTGGCCCGCCGTGCGC
GAGCGCATGCGCCGCACCGAGCCCGCCGCCGAGGGCGTGGGCGCCGCCAGC
CAGGACCTGGACCGCCACGGCGCCCTGACCAGCAGCAACACCCCCGCCACCA
ACGAGGCCTGCGCCTGGCTGCAGGCCCAGGAGGAGGACGGCGACGTGGGCT
TCCCCGTGCGCCCCCAGGTGCCCCTGCGCCCCATGACCTACAAGAGCGCCGT
GGACCTGAGCTTCTTCCTGAAGGAGAAGGGCGGCCTGGAGGGCCTGATCTAC
AGCCGCAAGCGCCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGCT
TCTTCCCCGGCTGGCAGAACTACACCAGCGGCCCCGGCGTGCGCTTCCCCCTG
ACCTTCGGCTGGTGCTTCAAGCTGGTGCCCGTGGACCCCCGCGAGGTGAAGG
AGGCCAACGAGGGCGAGGACAACTGCCTGCTGCACCCCATGAGCCAGCACG
GCGCCGAGGACGAGGACCGCGAGGTGCTGAAGTGGAAGTTCGACAGCCTGC
TGGCCCACCGCCACATGGCCCGCGAGCTGCACCCCGAGTACTACAAGGACTG
CTGA

FIGURE 67

ATGCGCGCCCGCGGCATCCTGAAGAACTACCGCCACTGGTGGATCTGGGGCATCCT
GGGCTTCTGGATGCTGATGATGTGCAACGTGAAGGGCCTGTGGGTGACCGTGTACTA
CGGCGTGCCCGTGGGCCGCGAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCA
AGGCCTACGAGAAGGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCCACC
GACCCCAACCCCCAGGAGGTGATCCTGGGCAACGTGACCGAGAACTTCAACATGTG
GAAGAACGACATGGTGGACCAGATGCAGGAGGACATCATCAGCCTGTGGGACCAGA
GCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTGACCCTGAACTGCACCAACG
CCACCGTGAACTACAACAACACCAGCAAGGACATGAAGAACTGCAGCTTCTACGTG
ACCACCGAGCTGCGCGACAAGAAGAAGAAGGAGAACGCCCTGTTCTACCGCCTGGA
CATCGTGCCCCTGAACAACCGCAAGAACGGCAACATCAACAACTACCGCCTGATCA
ACTGCAACACCAGCGCCATCACCCAGGCCTGCCCCAAGGTGAGCTTCGACCCCATCC
CCATCCACTACTGCGCCCCCGCCGGCTACGCCCCCCTGAAGTGCAACAACAAGAAG
TTCAACGGCATCGGCCCCTGCGACAACGTGAGCACCGTGCAGTGCACCCACGGCAT
CAAGCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGA
TCATCATCCGCAGCGAGAACCTGACCAACAACGTGAAGACCATCATCGTGCACCTG
AACGAGAGCATCGAGATCAAGTGCACCCGCCCCGGCAACAACACCCGCAAGAGCGT
GCGCATCGGCCCCGGCCAGGCCTTCTACGCCACCGGCGACATCATCGGCGACATCC
GCCAGGCCCACTGCAACATCAGCAAGAACGAGTGGAACACCACCCTGCAGCGCGTG
AGCCAGAAGCTGCAGGAGCTGTTCCCCAACAGCACCGGCATCAAGTTCGCCCCCCA
CAGCGGCGGCGACCTGGAGATCACCACCCACAGCTTCAACTGCGGCGGCGAGTTCT
TCTACTGCAACACCACCGACCTGTTCAACAGCACCTACAGCAACGGCACCTGCACCA
ACGGCACCTGCATGAGCAACAACACCGAGCGCATCACCCTGCAGTGCCGCATCAAG
CAGATCATCAACATGTGGCAGGAGGTGGGCCGCGCCATGTACGCCCCCCCCATCGC
CGGCAACATCACCTGCCGCAGCAACATCACCGGCCTGCTGCTGACCCGCGACGGCG
GCGACAACAACACCGAGACCGAGACCTTCCGCCCCGGCGGCGGCGACATGCGCGAC
AACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGGAGATCAAGCCCCTGGGCGT
GGCCCCCACCGCCGCCAAGCGCCGCGTGGTGGAGCGCGAGAAGCGCGCCGTGGGCA
TCGGCGCCGTGTTCCTGGGCTTCCTGGGCGCCGCCGGCAGCACCATGGGCGCCGCCA
GCATCACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGCAGCAGCAG
AGCAACCTGCTGCGCGCCATCGAGGCCCAGCAGCACATGCTGCAGCTGACCGTGTG
GGGCATCAAGCAGCTGCAGGCCCGCGTGCTGGCCATCGAGCGCTACCTGCAGGACC
AGCAGCTGCTGGGCCTGTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCAACGTG
CTGTGGAACAGCAGCTGGAGCAACAAGACCCAGAGCGACATCTGGGACAACATGAC
CTGGATGCAGTGGGACCGCGAGATCAGCAACTACACCAACACCATCTACCGCCTGC
TGGAGGACAGCCAGAGCCAGCAGGAGCGCAACGAGAAGGACCTGCTGGCCCTGGA
CCGCTGGAACAACCTGTGGAACTGGTTCAGCATCACCAACTGGCTGTGGTACATCAA
GATCTTCATCATGATCGTGGGCGGCCTGATCGGCCTGCGCATCATCTTCGCCGTGCT
GAGCCTGGTGAACCGCGTGCGCCAGGGCTACAGCCCCCTGAGCCTGCAGACCCTGA
TCCCCAACCCCCCGCGGCCCCGACCGCCTGGGCGGCATCGAGGAGGAGGGCGGCGAG
CAGGACAGCAGCCGCAGCATCCGCCTGGTGAGCGGCTTCCTGACCCTGGCCTGGGA
CGACCTGCGCAGCCTGTGCCTGTTCTGCTACCACCGCCTGCGCGACTTCATCCTGAT
CGTGGTGCGCGCCGTGGAGCTGCTGGGCCACAGCAGCCTGCGCGGCCTGCAGCGCG
GCTGGGGCACCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGCCTGGAGCTG
AAGAAGAGCGCCATCAACCTGCTGGACACCATCGCCATCGCCGTGGCCGAGGGCAC
CGACCGCATCCTGGAGTTCATCCAGAACCTGTGCCGCGGCATCCGCAACGTGCCCCG
CCGCATCCGCCAGGGCTTCGAGGCCGCCCTGCAGTAA

FIGURE 68

ATGAGAGCGAGGGGGATACTGAAGAATTATCGACACTGGTGGATATGGGGCATCTT
AGGCTTTTGGATGCTAATGATGTGTAATGTGAAGGGCTTGTGGGTCACAGTCTACTA
CGGGGTACCTGTGGGGAGAGAAGCAAAAACTACTCTATTTTGTGCATCAGATGCTA
AAGCATATGAGAAAGAAGTGCATAATGTCTGGGCTACACATGCCTGTGTACCCACA
GACCCCAACCCACAAGAAGTGATTTTGGGCAATGTAACAGAAAATTTTAACATGTG
GAAAAATGACATGGTGGATCAGATGCAGGAAGATATAATCAGTTTATGGGATCAAA
GCCTTAAGCCATGTGTAAAATTGACCCCACTCTGTGTCACTTTAAACTGTACAAATG
CAACTGTTAACTACAATAATACCTCTAAAGACATGAAAAATTGCTCTTTCTATGTAA
CCACAGAATTAAGAGATAAGAAAAAGAAAGAAAATGCACTTTTTTATAGACTTGAT
ATAGTACCACTTAATAATAGGAAGAATGGGAATATTAACAACTATAGATTAATAAA
TTGTAATACCTCAGCCATAACACAAGCCTGTCCAAAAGTCTCGTTTGACCCAATTCC
TATACATTATTGTGCTCCAGCTGGTTATGCGCCTCTAAAATGTAATAATAAGAAATT
CAATGGAATAGGACCATGCGATAATGTCAGCACAGTACAATGTACACATGGAATTA
AGCCAGTGGTATCAACTCAATTACTGTTAAATGGTAGCCTAGCAGAAGAAGAGATA
ATAATTAGATCTGAAAATCTGACAAACAATGTCAAAACAATAATAGTACATCTTAAT
GAATCTATAGAGATTAAATGTACAAGACCTGGCAATAATACAAGAAAGAGTGTGAG
AATAGGACCAGGACAAGCATTCTATGCAACAGGAGACATAATAGGAGATATAAGAC
AAGCACATTGTAACATTAGTAAAAATGAATGGAATACAACTTTACAAAGGGTAAGT
CAAAAATTACAAGAACTCTTCCCTAATAGTACAGGGATAAAATTTGCACCACACTCA
GGAGGGGACCTAGAAATTACTACACATAGCTTTAATTGTGGAGGAGAATTTTTCTAT
TGCAATACAACAGACCTGTTTAATAGTACATACAGTAATGGTACATGCACTAATGGT
ACATGCATGTCTAATAATACAGAGCGCATCACACTCCAATGCAGAATAAAACAAAT
TATAAACATGTGGCAGGAGGTAGGACGAGCAATGTATGCCCCTCCCATTGCAGGAA
ACATAACATGTAGATCAAATATTACAGGACTACTATTAACACGTGATGGAGGAGAT
AATAATACTGAAACAGAGACATTCAGACCTGGAGGAGGAGACATGAGGGACAATTG
GAGAAGTGAATTATATAAATACAAGGTGGTAGAAATTAAACCATTAGGAGTAGCAC
CCACTGCTGCAAAAAGGAGAGTGGTGGAGAGAGAAAAAAGAGCAGTAGGAATAGG
AGCTGTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCATCAAT
AACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAACAGCAAAGTA
ATTTGCTGAGGGCTATAGAGGCGCAACAGCATATGTTGCAACTCACGGTCTGGGGC
ATTAAGCAGCTCCAGGCAAGAGTCCTGGCTATAGAGAGATACCTACAGGATCAACA
GCTCCTAGGACTGTGGGGCTGCTCTGGAAAACTCATCTGCACCACTAATGTGCTTTG
GAACTCTAGTTGGAGTAATAAAACTCAAAGTGATATTTGGGATAACATGACCTGGAT
GCAGTGGGATAGGGAAATTAGTAATTACACAAACACAATATACAGGTTGCTTGAAG
ACTCGCAAAGCCAGCAGGAAAGAAATGAAAAAGATTTACTAGCATTGGACAGGTGG
AACAATCTGTGGAATTGGTTTAGCATAACAAATTGGCTGTGGTATATAAAAATATTC
ATAATGATAGTAGGAGGCTTGATAGGTTTAAGAATAATTTTTGCTGTGCTCTCTCTA
GTAAATAGAGTTAGGCAGGGATACTCACCCTTGTCATTGCAGACCCTTATCCCAAAC
CCGAGGGGACCCGACAGGCTCGGAGGAATCGAAGAAGAAGGTGGAGAGCAAGACA
GCAGCAGATCCATTCGATTAGTGAGCGGATTCTTGACACTTGCCTGGGACGACCTAC
GAAGCCTGTGCCTCTTCTGCTACCACCGATTGAGAGACTTCATATTAATTGTAGTGA
GAGCAGTGGAACTTCTGGGACACAGTAGTCTCAGGGGACTGCAGAGGGGGTGGGGA
ACCCTTAAGTATTTGGGGAGTCTTGTGCAATATTGGGGTCTAGAGTTAAAAAAGAGT
GCTATTAATCTGCTTGATACTATAGCAATAGCAGTAGCTGAAGGAACAGATAGGATT
CTAGAATTCATACAAAACCTTTGTAGAGGTATCCGCAACGTACCTAGAAGAATAAG
ACAGGGCTTCGAAGCAGCTTTGCAATAA

FIGURE 69

Gag_TV2_C_ZAopt (SEQ ID NO:99)

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGGCAAGCTGGACAAGTGGGAG
AAGATCCGCCTGCGCCCCGGCGGCCGCAAGCACTACATGCTGAAGCACCTGG
TGTGGGCCAGCCGCGAGCTGGAGCGCTTCGCCGTGAACCCCGGCCTGCTGGA
GACCAGCGACGGCTGCCGCCAGATCATCAAGCAGCTGCAGCCCGCCCTGCAG
ACCGGCACCGAGGAGATCCGCAGCCTGTTCAACACCGTGGCCACCCTGTACT
GCGTGCACAAGGGCATCGACGTGCGCGACACCAAGGAGGCCCTGGACAAGA
TCGAGGAGGAGCAGAACAAGTGCCAGCAGAAGACCCAGCAGGCCGAGGCCG
CCGACAAGAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCC
AGATGGTGCACCAGGCCATCAGCCCCCGCACCCTGAACGCCTGGGTGAAGGT
GATCGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTCACCGCCCTG
AGCGAGGGCGCCACCCCCCAGGACCTGAACACCATGCTGAACACCGTGGGC
GGCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAACGAGGAGGCC
GCCGAGTGGGACCGCCTGCACCCCGTGCACGCCGGCCCCGTGGCCCCCGGCC
AGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCACCACCAGCACCCTGCA
GGAGCAGATCGCCTGGATGACCAGCAACCCCCCCATCCCCGTGGGCGACATC
TACAAGCGCTGGATCATCCTGGGCCTGAACAAGATCGTGCGCATGTACAGCC
CCGTGAGCATCCTGGACATCAAGCAGGGCCCCAAGGAGCCCTTCCGCGACTA
CGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAGAGCACCCAGGAGGTG
AAGAACTGGATGACCGACACCCTGCTGGTGCAGAACGCCAACCCCGACTGCA
AGACCATCCTGCGCGCCCTGGGCCCCGGCGCCAGCCTGGAGGAGATGATGAC
CGCCTGCCAGGGCGTGGGCGGCCCCAGCCACAAGGCCCGCGTGCTGGCCGAG
GCCATGAGCCAGGCCAACAACACCAGCGTGATGATCCAGAAGAGCAACTTC
AAGGGCCCCCGCCGCGCCGTGAAGTGCTTCAACTGCGGCCGCGAGGGCCACA
TCGCCCGCAACTGCCGCGCCCCCCGCAAGCGCGGCTGCTGGAAGTGCGGCAA
GGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCCTGGG
CAAGATCTGGCCCAGCCACAAGGGCCGCCCCGGCAACTTCCTGCAGAGCCGC
CCCGAGCCCACCGCCCCCCCCCTGGAGCCCACCGCCCCCCCCGCCGAGAGCT
TCAAGTTCAAGGAGACCCCCAAGCAGGAGCCCAAGGACCGCGAGCCCCTGA
CCAGCCTGAAGAGCCTGTTCGGCAGCGACCCCCTGAGCCAGTAA

FIGURE 70

Gag_TV2_C_ZAwt (SEQ ID NO:100)

ATGGGTGCGAGAGCGTCAATATTAAGAGGGGGAAAATTAGACAAATGGGAA
AAAATTAGGTTACGGCCAGGGGGGAGAAAACACTATATGCTAAAACACCTA
GTATGGGCAAGCAGAGAGCTGGAAAGATTTGCAGTTAACCCTGGCCTTTTAG
AGACATCAGACGGATGTAGACAAATAATAAAACAGCTACAACCAGCTCTTCA
GACAGGAACAGAGGAAATTAGATCATTATTTAACACAGTAGCAACTCTCTAT
TGTGTACATAAAGGGATAGATGTACGAGACACCAAGGAAGCCTTAGACAAG
ATAGAGGAGGAACAAAACAAATGTCAGCAAAAAACACAGCAGGCGGAAGCG
GCTGACAAAAAGGTCAGTCAAAATTATCCTATAGTGCAGAACCTCCAAGGGC
AAATGGTACACCAGGCCATATCACCTAGAACCTTGAATGCATGGGTAAAAGT
AATAGAGGAGAAGGCTTTTAGCCCAGAGGTAATACCCATGTTTACAGCATTA
TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGTTAAATACAGTGGGGG
GACATCAAGCAGCCATGCAAATGTTAAAAGATACCATCAATGAGGAGGCTGC
AGAATGGGATAGGTTACATCCAGTACATGCAGGGCCTGTTGCACCAGGCCAG
ATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTTCAA
GAACAAATAGCATGGATGACAAGTAACCCACCTATCCCAGTAGGGGACATCT
ATAAAAGGTGGATAATTCTGGGGTTAAATAAAATAGTAAGAATGTACAGCCC
TGTCAGCATTTTAGACATAAAACAAGGACCAAAGGAACCCTTTAGAGACTAT
GTAGACCGGTTCTTCAAAACTTTAAGAGCTGAACAATCTACACAAGAGGTAA
AAAATTGGATGACAGACACCTTGTTAGTCCAAAATGCGAACCCAGATTGTAA
GACCATTTTAAGAGCATTAGGACCAGGGGCTTCATTAGAAGAAATGATGACA
GCATGTCAGGGAGTGGGAGGACCTAGCCACAAAGCAAGAGTTTTGGCTGAG
GCAATGAGCCAAGCAAACAATACAAGTGTAATGATACAGAAAAGCAATTTTA
AAGGCCCTAGAAGAGCTGTTAAATGTTTCAACTGTGGCAGGGAAGGGCACAT
AGCCAGGAATTGCAGGGCCCCTAGGAAAAGGGGCTGTTGGAAATGTGGAAA
GGAAGGACACCAAATGAAAGACTGTACTGAGAGGCAGGCTAATTTTTTAGGG
AAAATTTGGCCTTCCCACAAGGGGAGGCCAGGGAATTTCCTTCAGAGCAGAC
CAGAGCCAACAGCCCCACCACTAGAACCAACAGCCCCACCAGCAGAGAGCT
TCAAGTTCAAGGAGACTCCGAAGCAGGAGCCGAAAGACAGGGAACCTTTAA
CTTCCCTCAAATCACTCTTTGGCAGCGACCCCTTGTCTCAATAA

FIGURE 71

Nef_TV2_C_ZAopt (SEQ ID NO:101)

ATGGGCGGCAAGTGGAGCAAGAGCAGCATCATCGGCTGGCCCGAGGTGCGC
GAGCGCATCCGCCGCACCCGCAGCGCCGCCGAGGGCGTGGGCAGCGCCAGC
CAGGACCTGGAGAAGCACGGCGCCCTGACCACCAGCAACACCGCCCACAAC
AACGCCGCCTGCGCCTGGCTGGAGGCCCAGGAGGAGGAGGGCGAGGTGGGC
TTCCCCGTGCGCCCCCAGGTGCCCCTGCGCCCCATGACCTACAAGGCCGCCAT
CGACCTGAGCTTCTTCCTGAAGGAGAAGGGCGGCCTGGAGGGCCTGATÇTAC
AGCAAGAAGCGCCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGC
TTCTTCCCCGACTGGCAGAACTACACCCCCGGCCCCGGCGTGCGCTTCCCCCT
GACCTTCGGCTGGTACTTCAAGCTGGAGCCCGTGGACCCCCGCGAGGTGGAG
GAGGCCAACGAGGGCGAGAACAACTGCCTGCTGCACCCCATGAGCCAGCAC
GGCATGGAGGACGAGGACCGCGAGGTGCTGCGCTGGAAGTTCGACAGCACC
CTGGCCCGCCGCCACATGGCCCGCGAGCTGCACCCCGAGTACTACAAGGACT
GCTGA

FIGURE 72

Nef_TV2_C_ZAwt (SEQ ID NO:102)

ATGGGGGGCAAGTGGTCAAAAAGCAGTATAATTGGATGGCCTGAAGTAAGA
GAAAGAATCAGACGAACTAGGTCAGCAGCAGAGGGAGTAGGATCAGCGTCT
CAAGACTTAGAGAAACATGGGGCACTTACAACCAGCAACACAGCCCACAAC
AATGCTGCTTGCGCCTGGCTGGAAGCGCAAGAGGAGGAAGGAGAAGTAGGC
TTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATGACTTATAAAGCAGCAAT
AGATCTCAGCTTCTTTTTAAAAGAAAAGGGGGGACTGGAAGGGTTAATTTAC
TCCAAGAAAAGGCAAGAGATCCTTGATTTGTGGGTTTATAACACACAAGGCT
TCTTCCCTGATTGGCAAAACTACACACCGGGACCAGGGGTCAGATTTCCACT
GACCTTTGGATGGTACTTCAAGCTAGAGCCAGTCGATCCAAGGGAAGTAGAA
GAGGCCAATGAAGGAGAAAACAACTGTTTACTACACCCTATGAGCCAGCATG
GAATGGAGGATGAAGACAGAGAAGTATTAAGATGGAAGTTTGACAGTACGC
TAGCACGCAGACACATGGCCCGCGAGCTACATCCGGAGTATTACAAAGACTG
CTGA

FIGURE 73

Pol_TV2_C_ZAopt (SEQ ID NO:103)

TTCTTCCGCGAGAACCTGGCCTTCCCCCAGGGCGAGGCCCGCGAGTTCCCCAGCGAGCAGACC
CGCGCCAACAGCCCCACCACCCGCACCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCAGGG
CGACAGCGAGGCCGGCGCCGAGCGCCAGGGCACCTTCAACTTCCCCCAGATCACCCTGTGGC
AGCGCCCCCTGGTGAGCATCAAGGTGGCCGGCCAGACCAAGGAGGCCCTGCTGGACACCGGC
GCCGACGACACCGTGCTGGAGGAGATCAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCGG
CGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCA
AGCGCGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGC
TGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGC
TGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAG
GCCCTGACCGAGATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCG
AGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG
CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCAT
CCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT
ACTTCAGCGTGCCCCTGGACGAGAGCTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA
ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC
CCCGCCATCTTCCAGAGCAGCATGACCCGCATCCTGGAGCCCTTCCGCACCCAGAACCCCGAG
GTGGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCA
CCGCGCCAAGATCGAGGAGCTGCGCGGCCACCTGCTGAAGTGGGGCTTCACCACCCCCGACA
AGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGG
ACCGTGCAGCCCATCCAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCT
GGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA
AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTG
GAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCGTGTACTACGACCCCAG
CAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCAACGACCAGTGGACCTACCAGATCTACC
AGGAGCCCTTCAAGAACCTGCGCACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAAC
GACGTGAAGCAGCTGGCCGAGGCCGTGCAGAAGATCACCCAGGAGAGCATCGTGATCTGGGG
CAAGACCCCCAAGTTCCGCCTGCCCATCCCCAAGGAGACCTGGGAGACCTGGTGGAGCGACT
ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGT
GGTACCAGCTGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCC
AACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACAAGGGCCGCCAGAAGGTGG
TGAGCTTCACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAG
GACAGCGGCCCCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGC
CCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAG
GAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGA
CAAGCTGGTGAGCAGCGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGG
AGGAGCACGAGAAGTACCACAGCAACTGGCGCGCCATGGCCAGCGAGTTCAACCTGCCCCCC
ATCGTGGCCAAGGAGATCGTGGCCAGCTGCGACAAGTGCCAGCTGAAGGGCGAGGCCATGCA
CGGCCAGGTGGACTGCAGCCCCGGCATCTGGCAGCTGGACTGCACCCACCTGGAGGGCAAGA
TCATCCTGGTGGCCGTGCACGTGGCCAGCGGCTACATGGAGGCCGAGGTGATCCCCGCCGAG
ACCGGCCAGGAGACCGCCTACTTCATCCTGAAGCTGGCCGGCCGCTGGCCCGTGAAGGTGATC
CACACCGACAACGGCAGCAACTTCACCAGCACCGCCGTGAAGGCCGCCTGCTGGTGGGCCGA
CATCCAGCGCGAGTTCGGCATCCCCTACAACCCCCAGAGCCAGGGCGTGGTGGAGAGCATGA
ACAAGGAGCTGAAGAAGATCATCGGCCAGGTGCGCGACCAGGCCGAGCACCTGAAGACCGCC
GTGCAGATGGCCGTGTTCATCCACAACTTCAAGCGCAAGGGCGGCATCGGCGGCTACAGCGC
CGGCGAGCGCATCATCGACATCATCGCCAGCGACATCCAGACCAAGGAGCTGCAGAAGCAGA
TCATCAAGATCCAGAACTTCCGCGTGTACTACCGCGACAGCCGCGACCCCATCTGGAAGGGCC
CCGCCAAGCTGCTGTGGAAGGGCGAGGGCGCCGTGGTGATCCAGGACAACAGCGACATCAAG
GTGGTGCCCCGCCGCAAGGCCAAGATCATCAAGGACTACGGCAAGCAGATGGCCGGCGCCGA
CTGCGTGGCCGGCCGCCAGGACGAGGAC

FIGURE 74

Pol_TV2_C_ZAwt (SEQ ID NO:104)

TTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGGAGGCCAGGGAATTTCCTTCAGAGCAGACC
AGAGCCAACAGCCCCACCACTAGAACCAACAGCCCCACCAGCAGAGAGCTTCAAGTTCAAGG
AGACTCCGAAGCAGGAGCCGAAAGACAGGGAACCTTTAACTTCCCTCAAATCACTCTTTGGCA
GCGACCCCTTGTCTCAATAAAAGTAGCGGGCCAAACAAAGGAGGCTCTTTTAGATACAGGAG
CAGATGATACAGTACTAGAAGAAATAAACTTGCCAGGAAAATGGAAACCAAAAATGATAGG
AGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAA
AAGGGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTT
GACTCAGCTTGGATGCACACTAAATTTTCCAATTAGCCCCATTGAAACTGTACCAGTAAAATT
AAAGCCAGGAATGGATGGCCCAAAGGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAA
GCATTAACAGAAATTTGTGAGGAAATGGAGAAGGAAGGAAAAATTACAAAAATTGGGCCTGA
AAATCCATATAACACTCCAGTATTTGCCATAAAGAAGAAGGACAGTACAAAGTGGAGAAAAT
TAGTAGATTTCAGGGAACTCAATAAAAGAACTCAAGACTTTTGGGAAGTCCAATTAGGAATA
CCACACCCAGCAGGGTTAAAAAAGAAAAAATCAGTGACAGTACTGGATGTGGGAGATGCATA
TTTTTCAGTCCCTTTAGATGAGAGCTTCAGAAAATATACTGCATTCACCATACCTAGTATAAAC
AATGAAACACCAGGGATTAGATATCAATATAATGTTCTTCCACAGGGATGGAAAGGATCACC
AGCAATATTCCAGAGTAGCATGACAAGAATCTTAGAGCCCTTTAGAACACAAAACCCAGAAG
TAGTTATCTATCAATATATGGATGACTTATATGTAGGATCTGACTTAGAAATAGGGCAACATA
GAGCAAAAATAGAGGAGTTAAGAGGACACCTATTGAAATGGGGATTTACCACACCAGACAAG
AAACATCAGAAAGAACCCCCATTTCTTTGGATGGGGTATGAACTCCATCCTGACAAATGGACA
GTACAGCCTATACAGCTGCCAGAAAAGGAGAGCTGGACTGTCAATGATATACAGAAGTTAGT
GGGAAAGTTAAACTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAACTGTGTAAAC
TCCTTAGGGGAGCCAAAGCACTAACAGACATAGTGCCACTGACTGAAGAAGCAGAATTAGAA
TTGGCTGAGAACAGGGAAATTCTAAAAGAACCAGTACATGGAGTATATTATGACCCATCAAA
AGATTTAATAGCTGAAATACAGAAACAGGGGAATGACCAATGGACATATCAAATTTACCAAG
AACCATTTAAAAATCTGAGAACAGGAAAGTATGCAAAAATGAGGACTGCCCACACTAATGAT
GTGAAACAGTTAGCAGAGGCAGTGCAAAAGATAACCCAGGAAAGCATAGTAATATGGGGAA
AAACTCCTAAATTTAGACTACCCATCCCAAAAGAAACATGGGAGACATGGTGGTCAGACTATT
GGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCCTAGTAAAATTGTGGT
ACCAGCTGGAAAAAGAACCCATAGTAGGGGCAGAAACTTTCTATGTAGATGGAGCAGCCAAT
AGGGAAACTAAAATAGGAAAAGCAGGGTATGTCACTGACAAAGGAAGGCAGAAAGTTGTTTC
CTTCACTGAAACAACAAATCAGAAGACTGAATTACAAGCAATTCAGCTAGCTTTGCAGGATTC
AGGGCCAGAAGTAAACATAGTAACAGACTCACAGTATGCATTAGGAATCATTCAAGCACAAC
CAGATAAGAGTGAATCAGAATTAGTCAGTCAAATAATAGAACAGTTGATAAAAAAGGAAAAA
GTCTACCTATCATGGGTACCAGCACATAAAGGAATTGGAGGAAATGAACAAGTAGACAAATT
AGTAAGTAGTGGAATCAGAAAAGTACTGTTTCTAGATGGAATAGATAAAGCTCAAGAAGAGC
ATGAAAAATATCACAGCAATTGGAGAGCAATGGCTAGTGAGTTTAATCTGCCACCCATAGTA
GCAAAGGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGGGAAGCCATGCATGGACA
AGTCGACTGTAGTCCAGGAATATGGCAATTAGACTGTACACATTTAGAAGGAAAAATCATCCT
AGTAGCAGTCCATGTAGCCAGTGGCTACATGGAAGCAGAGGTTATCCCAGCAGAAACAGGAC
AAGAAACAGCATACTTTATACTAAAATTAGCAGGAAGATGGCCAGTCAAAGTAATACATACA
GATAATGGCAGTAATTTCACCAGTACCGCAGTTAAGGCAGCCTGTTGGTGGGCAGATATCCAA
CGGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCCATGAATAAAGA
ATTAAAGAAAATCATAGGGCAAGTAAGAGATCAAGCTGAGCACCTTAAGACAGCAGTACAAA
TGGCAGTATTCATTCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAG
AGAATAATAGACATAATAGCATCAGACATACAAACTAAAGAATTACAAAAACAAATTATAAA
AATTCAAAATTTTCGGGTTTATTACAGAGACAGCAGAGACCCTATTTGGAAAGGACCAGCCAA
ACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGATATAAAGGTAGTAC
CAAGAAGGAAAGCAAAAATCATTAAGGACTATGGAAAACAGATGGCAGGTGCTGATTGTGTG
GCAGGTAGACAGGATGAAGAT

FIGURE 75

RevExon1_TV2_C_ZAopt (SEQ ID NO:105)

ATGGCCGGCCGCAGCGGCGACAGCGACGAGGCCCTGCTGCAGGCCATCAAG
ATCATCAAGATCCTGTACCAGAGC

FIGURE 76

RevExon1_TV2_C_ZAwt (SEQ ID NO:106)

ATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGCTCCTCCAAGCAATAAAG
ATCATCAAGATCCTCTACCAAAGCA

FIGURE 77

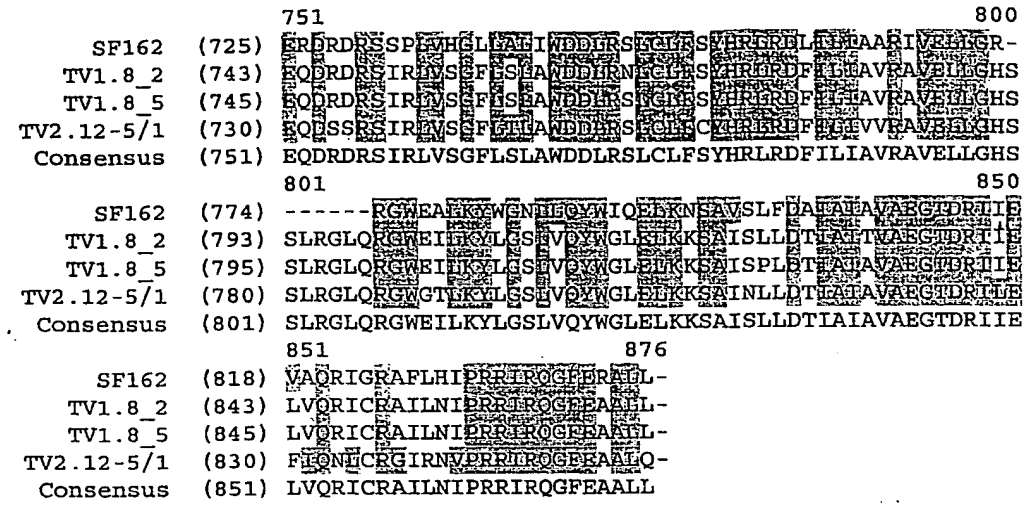


FIGURE 105C

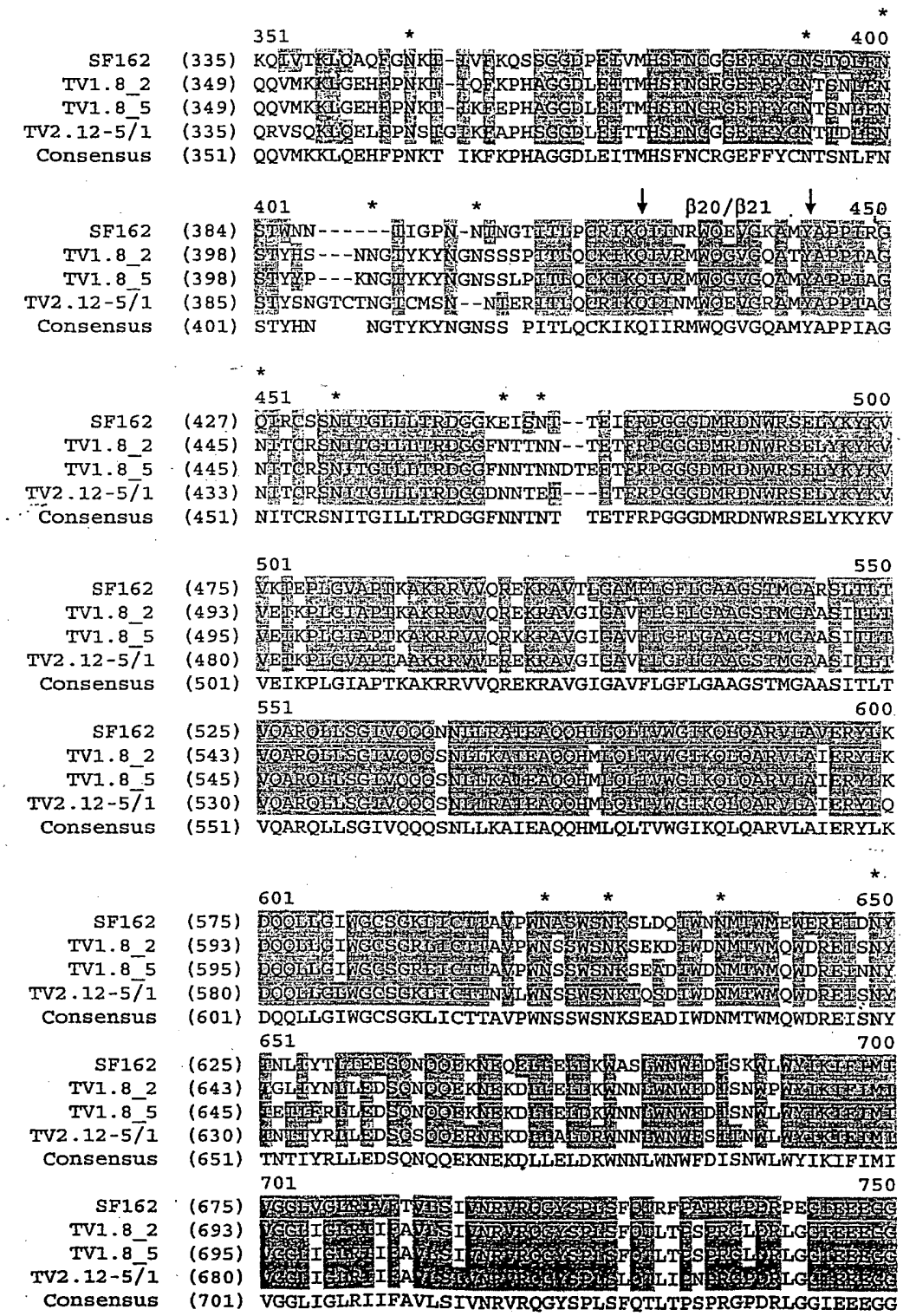


FIGURE 105B

↓: is the regions for β-sheet deletions

*: is the N-linked glycosylation sites for subtype C TV1 and TV2. Possible mutation (N→ Q) or deletions can be performed.

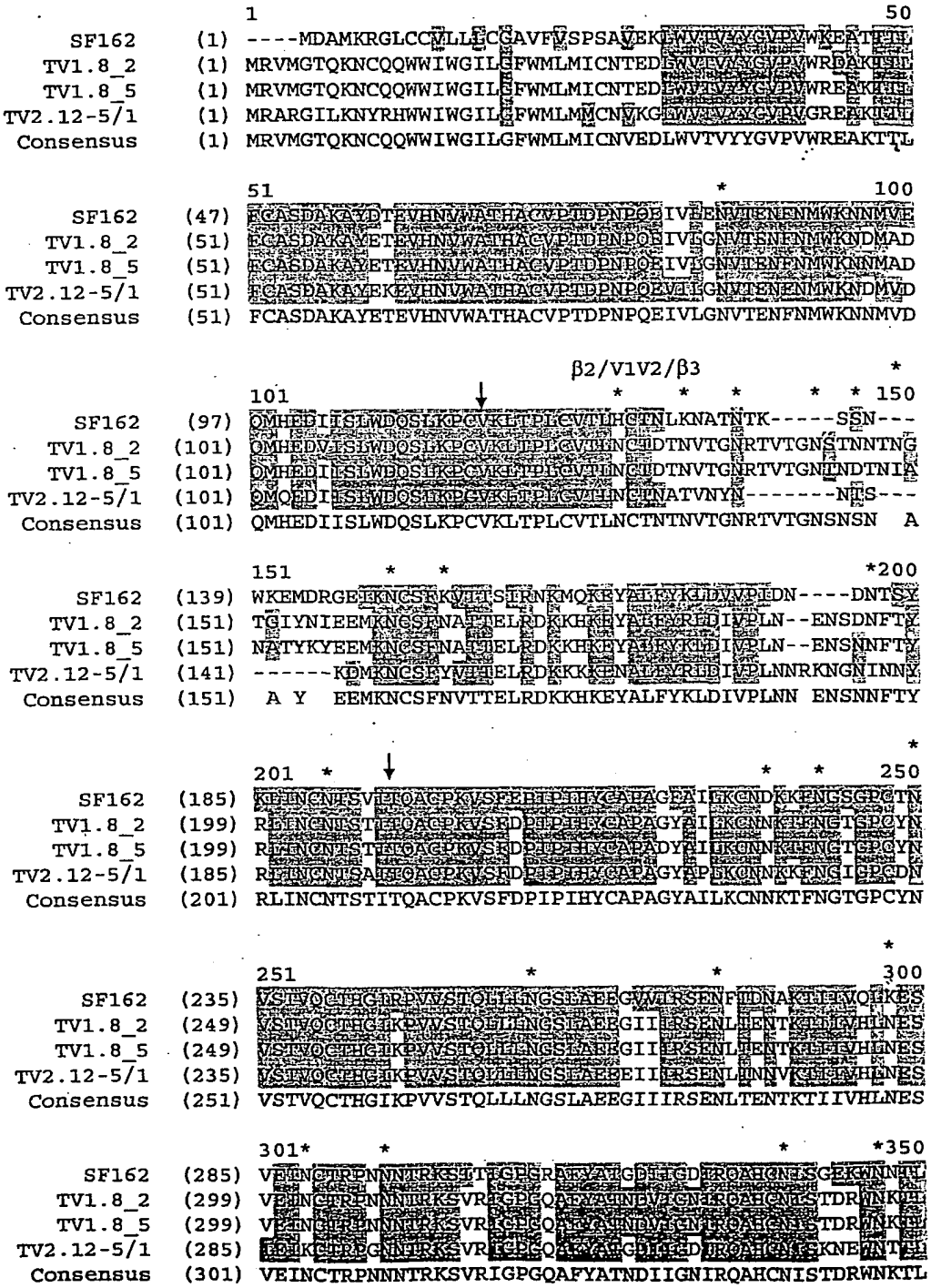


FIGURE 105A

NefD125G-Myr_TV2_C_ZAopt (SEQ ID NO:135)

ATGGCCGGCAAGTGGAGCAAGAGCAGCATCATCGGCTGGCCCGAGGTGCGC
GAGCGCATCCGCCGCACCCGCAGCGCCGCCGAGGGCGTGGGCAGCGCCAGC
CAGGACCTGGAGAAGCACGGCGCCCTGACCACCAGCAACACCGCCCACAAC
AACGCCGCCTGCGCCTGGCTGGAGGCCCAGGAGGAGGAGGGCGAGGTGGGC
TTCCCCGTGCGCCCCCAGGTGCCCCTGCGCCCCATGACCTACAAGGCCGCCAT
CGACCTGAGCTTCTTCCTGAAGGAGAAGGGCGGCCTGGAGGGCCTGATCTAC
AGCAAGAAGCGCCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGC
TTCTTCCCCCGGCTGGCAGAACTACACCCCCGGCCCCGGCGTGCGCTTCCCCCT
GACCTTCGGCTGGTACTTCAAGCTGGAGCCCGTGGACCCCCGCGAGGTGGAG
GAGGCCAACGAGGGCGAGAACAACTGCCTGCTGCACCCCATGAGCCAGCAC
GGCATGGAGGACGAGGACCGCGAGGTGCTGCGCTGGAAGTTCGACAGCACC
CTGGCCCGCCGCCACATGGCCCGCGAGCTGCACCCCGAGTACTACAAGGACT
GCTGA

FIGURE 104

NefD125G_TV2_C_ZAopt (SEQ ID NO:134)

ATGGGCGGCAAGTGGAGCAAGAGCAGCATCATCGGCTGGCCCGAGGTGCGC
GAGCGCATCCGCCGCACCCGCAGCGCCGCCGAGGGCGTGGGCAGCGCCAGC
CAGGACCTGGAGAAGCACGGCGCCCTGACCACCAGCAACACCGCCCACAAC
AACGCCGCCTGCGCCTGGCTGGAGGCCCAGGAGGAGGAGGGCGAGGTGGGC
TTCCCCGTGCGCCCCCAGGTGCCCCTGCGCCCCATGACCTACAAGGCCGCCAT
CGACCTGAGCTTCTTCCTGAAGGAGAAGGGCGGCCTGGAGGGCCTGATCTAC
AGCAAGAAGCGCCAGGAGATCCTGGACCTGTGGGTGTACAACACCCAGGGC
TTCTTCCCCGGCTGGCAGAACTACACCCCCGGCCCCGGCGTGCGCTTCCCCCT
GACCTTCGGCTGGTACTTCAAGCTGGAGCCCGTGGACCCCCGCGAGGTGGAG
GAGGCCAACGAGGGCGAGAACAACTGCCTGCTGCACCCCATGAGCCAGCAC
GGCATGGAGGACGAGGACCGCGAGGTGCTGCGCTGGAAGTTCGACAGCACC
CTGGCCCGCCGCCACATGGCCCGCGAGCTGCACCCCGAGTACTACAAGGACT
GCTGA

FIGURE 103

gp140mod.TV1.wtLnative (SEQ ID NO:133)

1 gaattcatga gagtgatggg gacacagaag aattgtcaac aatggtggat atggggcatc
61 ttaggcttct ggatgctaat gatttgtaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccaccacc gagctgcgcg acaagaagca caaggagtac
541 gccctgttct accgcctgga catcgtgccc ctgaacgaga acagcgacaa cttcacctac
601 cgcctgatca actgcaacac cagcaccatc acccaggcct gccccaaggt gagcttcgac
661 cccatcccca tccactactg cgcccccgcc ggctacgcca tcctgaagtg caacaacaag
721 accttcaacg gcaccggccc ctgctacaac gtgagcaccg tgcagtgcac ccacggcatc
781 aagcccgtgg tgagcaccca gctgctgctg aacggcagcc tggccgagga gggcatcatc
841 atccgcagcg agaacctgac cgagaacacc aagaccatca tcgtgcacct gaacgagagc
901 gtggagatca actgcacccg ccccaacaac aacacccgca agagcgtgcg catcggcccc
961 ggccaggcct tctacgccac caacgacgtg atcggcaaca tccgccaggc ccactgcaac
1021 atcagcaccg accgctggaa caagaccctg cagcaggtga tgaagaagct gggcgagcac
1081 ttccccaaca agaccatcca gttcaagccc cacgccggcg gcgacctgga gatcaccatg
1141 cacagcttca actgccgcgg cgagttcttc tactgcaaca ccagcaacct gttcaacagc
1201 acctaccaca gcaacaacgg cacctacaag tacaacggca acagcagcag ccccatcacc
1261 ctgcagtgca agatcaagca gatcgtgcgc atgtggcagg gcgtgggcca ggccacctac
1321 gccccccccca tcgccggcaa catcacctgc cgcagcaaca tcaccggcat cctgctgacc
1381 cgcgacggcg gcttcaacac caccaacaac accgagacct tccgccccgg cggcggcgac
1441 atgcgcgaca actggcgcag cgagctgtac aagtacaagg tggtggagat caagcccctg
1501 ggcatcgccc ccaccaaggc caagcgccgc gtggtgcagc gcgagaagcg cgccgtgggc
1561 atcggcgccg tgttcctggg cttcctgggc gccgccggca gcaccatggg cgccgccagc
1621 atcaccctga ccgtgcaggc ccgccagctg ctgagcggca tcgtgcagca gcagagcaac
1681 ctgctgaagg ccatcgaggc ccagcagcac atgctgcagc tgaccgtgtg gggcatcaag
1741 cagctgcagg cccgcgtgct ggccatcgag cgctacctga aggaccagca gctgctgggc
1801 atctggggct gcagcggccg cctgatctgc accaccgccg tgccctggaa cagcagctgg
1861 agcaacaaga gcgagaagga catctgggac aacatgacct ggatgcagtg ggaccgcgag
1921 atcagcaact acaccggcct gatctacaac ctgctggagg acagccagaa ccagcaggag
1981 aagaacgaga aggacctgct ggagctggac aagtggaaca acctgtggaa ctggttcgac
2041 atcagcaact ggccctggta catctaactc gag

FIGURE 102

gp140mod.TV1 (SEQ ID NO:132)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccaccacc gagctgcgcg acaagaagca caaggagtac
541 gccctgttct accgcctgga catcgtgccc ctgaacgaga acagcgacaa cttcacctac
601 cgcctgatca actgcaacac cagcaccatc acccaggcct gccccaaggt gagcttcgac
661 cccatccccа tccactactg cgcccccgcc ggctacgcca tcctgaagtg caacaacaag
721 accttcaacg gcaccggccc ctgctacaac gtgagcaccg tgcagtgcac ccacggcatc
781 aagcccgtgg tgagcaccca gctgctgctg aacggcagcc tggccgagga gggcatcatc
841 atccgcagcg agaacctgac cgagaacacc aagaccatca tcgtgcacct gaacgagagc
901 gtggagatca actgcacccg ccccaacaac aacacccgca agagcgtgcg catcggcccc
961 ggccaggcct tctacgccac caacgacgtg atcggcaaca tccgccaggc ccactgcaac
1021 atcagcaccg accgctggaa caagaccctg cagcaggtga tgaagaagct gggcgagcac
1081 ttccccaaca agaccatcca gttcaagccc cacgccggcg gcgacctgga gatcaccatg
1141 cacagcttca actgccgcgg cgagttcttc tactgcaaca ccagcaacct gttcaacagc
1201 acctaccaca gcaacaacgg cacctacaag tacaacggca acagcagcag ccccatcacc
1261 ctgcagtgca agatcaagca gatcgtgcgc atgtggcagg gcgtgggcca ggccacctac
1321 gccccccccа tcgccggcaa catcacctgc cgcagcaaca tcaccggcat cctgctgacc
1381 cgcgacggcg gcttcaacac caccaacaac accgagacct tccgccccgg cggcggcgac
1441 atgcgcgaca actggcgcag cgagctgtac aagtacaagg tggtggagat caagcccctg
1501 ggcatcgccc ccaccaaggc caagcgccgc gtggtgcagc gcgagaagcg cgccgtgggc
1561 atcggcgccg tgttcctggg cttcctgggc gccgccggca gcaccatggg cgccgccagc
1621 atcaccctga ccgtgcaggc ccgccagctg ctgagcggca tcgtgcagca gcagagcaac
1681 ctgctgaagg ccatcgaggc ccagcagcac atgctgcagc tgaccgtgtg gggcatcaag
1741 cagctgcagg cccgcgtgct ggccatcgag cgctacctga aggaccagca gctgctgggc
1801 atctggggct gcagcggccg cctgatctgc accaccgccg tgccctggaa cagcagctgg
1861 agcaacaaga gcgagaagga catctgggac aacatgacct ggatgcagtg ggaccgcgag
1921 atcagcaact acaccggcct gatctacaac ctgctggagg acagccagaa ccagcaggag
1981 aagaacgaga aggacctgct ggagctggac aagtggaaca acctgtggaa ctggttcgac
2041 atcagcaact ggccctggta catctaactc gag

FIGURE 101

gp140mod.TV1.tpa1 (SEQ ID NO:131)

1 atggatgcaa tgaagagagg gctctgctgt gtgctgctgc tgtgtggagc agtcttcgtt
61 tcgcccagcg ccagcaccga ggacctgtgg gtgaccgtgt actacggcgt gcccgtgtgg
121 cgcgacgcca agaccaccct gttctgcgcc agcgacgcca aggcctacga gaccgaggtg
181 cacaacgtgt gggccaccca cgcctgcgtg cccaccgacc ccaaccccca ggagatcgtg
241 ctgggcaacg tgaccgagaa cttcaacatg tggaagaacg acatggccga ccagatgcac
301 gaggacgtga tcagcctgtg ggaccagagc ctgaagccct gcgtgaagct gaccccctg
361 tgcgtgaccc tgaactgcac cgacaccaac gtgaccggca accgcaccgt gaccggcaac
421 agcaccaaca acaccaacgg caccggcatc tacaacatcg aggagatgaa gaactgcagc
481 ttcaacgcca ccaccgagct gcgcgacaag aagcacaagg agtacgccct gttctaccgc
541 ctggacatcg tgcccctgaa cgagaacagc gacaacttca cctaccgcct gatcaactgc
601 aacaccagca ccatcaccca ggcctgcccc aaggtgagct tcgaccccat ccccatccac
661 tactgcgccc ccgccggcta cgccatcctg aagtgcaaca acaagacctt caacggcacc
721 ggcccctgct acaacgtgag caccgtgcag tgcacccacg gcatcaagcc cgtggtgagc
781 acccagctgc tgctgaacgg cagcctggcc gaggagggca tcatcatccg cagcgagaac
841 ctgaccgaga acaccaagac catcatcgtg cacctgaacg agagcgtgga gatcaactgc
901 acccgcccca acaacaacac ccgcaagagc gtgcgcatcg gccccggcca ggccttctac
961 gccaccaacg acgtgatcgg caacatccgc caggcccact gcaacatcag caccgaccgc
1021 tggaacaaga ccctgcagca ggtgatgaag aagctgggcg agcacttccc caacaagacc
1081 atccagttca agccccacgc cggcggcgac ctggagatca ccatgcacag cttcaactgc
1141 cgcggcgagt tcttctactg caacaccagc aacctgttca acagcaccta ccacagcaac
1201 aacggcacct acaagtacaa cggcaacagc agcagcccca tcaccctgca gtgcaagatc
1261 aagcagatcg tgcgcatgtg gcagggcgtg ggccaggcca cctacgcccc ccccatcgcc
1321 ggcaacatca cctgccgcag caacatcacc ggcatcctgc tgacccgcga cggcggcttc
1381 aacaccacca acaacaccga gaccttccgc cccggcggcg gcgacatgcg cgacaactgg
1441 cgcagcgagc tgtacaagta caaggtggtg gagatcaagc ccctgggcat cgcccccacc
1501 aaggccaagc gccgcgtggt gcagcgcgag aagcgcgccg tgggcatcgg cgccgtgttc
1561 ctgggcttcc tgggcgccgc cggcagcacc atgggcgccg ccagcatcac cctgaccgtg
1621 caggcccgcc agctgctgag cggcatcgtg cagcagcaga gcaacctgct gaaggccatc
1681 gaggcccagc agcacatgct gcagctgacc gtgtggggca tcaagcagct gcaggcccgc
1741 gtgctggcca tcgagcgcta cctgaaggac cagcagctgc tgggcatctg gggctgcagc
1801 ggccgcctga tctgcaccac cgccgtgccc tggaacagca gctggagcaa caagagcgag
1861 aaggacatct gggacaacat gacctggatg cagtgggacc gcgagatcag caactacacc
1921 ggcctgatct acaacctgct ggaggacagc cagaaccagc aggagaagaa cgagaaggac
1981 ctgctggagc tggacaagtg gaacaacctg tggaactggt tcgacatcag caactggccc
2041 tggtacatct aa

FIGURE 100

1 atgagagtga tggggacaca gaagaattgt caacaatggt ggatatgggg catcttaggc
61 ttctggatgc taatgatttg taacacggag gacttgtggg tcacagtcta ctatggggta
121 cctgtgtgga gagacgcaaa aactactcta ttctgtgcat cagatgctaa agcatatgag
181 acagaagtgc ataatgtctg ggctacacat gcctgtgtac ccacagaccc caacccacaa
241 gaaatagttt tgggaaatgt aacagaaaat tttaatatgt ggaaaaatga catggcagat
301 cagatgcatg aggatgtaat cagtttatgg gatcaaagcc taaagccatg tgtaaagttg
361 accccactct gtgtcacttt aaactgtaca gatacaaatg ttacaggtaa tagaactgtt
421 acaggtaata gtaccaataa tacaaatggt acaggtattt ataacattga agaaatgaaa
481 aattgctctt tcaatgcaac cacagaatta agagataaga aacataaaga gtatgcactc
541 ttttatagac ttgatatagt accacttaat gagaatagtg acaactttac atatagatta
601 ataaattgca atacctcaac cataacacaa gcctgtccaa aggtctcttt tgacccgatt
661 cctatacatt actgtgctcc agctggttat gcgattctaa agtgtaataa taagacattc
721 aatgggacag gaccatgtta taatgtcagc acagtacaat gtacacatgg aattaagcca
781 gtggtatcaa ctcaattact gttaaatggt agtctagcag aagaagggat aataattaga
841 tctgaaaatt tgacagagaa taccaaaaca ataatagtac accttaatga atctgtagag
901 attaattgta caagacccaa caataataca agaaaaagtg taaggatagg accaggacaa
961 gcattctatg caacaaatga tgtaatagga aacataagac aagcacattg taacattagt
1021 acagatagat ggaacaaaac tttacaacag gtaatgaaaa aattaggaga gcatttccct
1081 aataaaacaa tacaatttaa accacatgca ggaggggatc tagaaattac aatgcatagc
1141 tttaattgta gaggagaatt tttctattgt aatacatcaa acctgtttaa tagcacatac
1201 cactctaata atggtacata caaatacaat ggtaattcaa gctcacccat cacactccaa
1261 tgtaaaataa aacaaattgt acgcatgtgg caaggggtag gacaagcaac gtatgcccct
1321 cccattgcag gaaacataac atgtagatca aacatcacag gaatactatt gacacgtgat
1381 ggaggattta acaccacaaa caacacagag acattcagac ctggaggagg agatatgagg
1441 gataactgga gaagtgaatt atataaatat aaagtagtag aaattaagcc attgggaata
1501 gcacccacta aggcaaaaag aagagtggtg cagagagaaa aaagagcagt gggaatagga
1561 gctgtgttcc ttgggttctt gggagcagca ggaagcacta tgggcgcagc gtcaataacg
1621 ctgacggtac aggccagaca actgttgtct ggtatagtgc aacagcaaag caatttgctg
1681 aaggctatag aggcgcaaca gcatatgttg caactcacag tctggggcat taagcagctc
1741 caggcgagag tcctggctat agaaagatac ctaaaggatc aacagctcct agggatttgg
1801 ggctgctctg gaagactcat ctgcaccact gctgtgcctt ggaactccag ttggagtaat
1861 aaatctgaaa aagatatttg ggataacatg acttggatgc agtgggatag agaaattagt
1921 aattacacag gcttaatata caatttgctt gaagactcgc aaaaccagca ggaaaagaat
1981 gaaaaagatt tattagaatt ggacaagtgg aacaatctgt ggaattggtt tgacatatca
2041 aactggccgt ggtatataaa aatattcata atgatagtag gaggcttgat aggtttaaga
2101 ataatttttg ctgtgctttc tatagtgaat agagttaggc agggatactc acctttgtca
2161 tttcagaccc ttaccccaag cccgagggga ctcgacaggc tcggaggaat cgaagaagaa
2221 ggtggagagc aagacagaga cagatccata cgattggtga gcggattctt gtcgcttgcc
2281 tgggacgatc tgcggaacct gtgcctcttc agctaccacc gcttgagaga cttcatatta
2341 attgcagtga gggcagtgga acttctggga cacagcagtc tcaggggact acagaggggg
2401 tgggaaatcc ttaagtatct gggaagtctt gtgcaatatt ggggtctaga gctaaaaaag
2461 agtgctatta gtctgcttga taccatagca ataacagtag ctgaaggaac agataggatt
2521 atagaattag tacaaagaat ttgtagagct atcctcaaca tacctagaag aataagacag
2581 ggctttgaag cagctttgct ataa

FIGURE 99

1 gaattcatga gagtgatggg gacacagaag aattgtcaac aatggtggat atggggcatc
61 ttaggcttct ggatgctaat gatttgtaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccaccacc gagctgcgcg acaagaagca caaggagtac
541 gccctgttct accgcctgga catcgtgccc ctgaacgaga acagcgacaa cttcacctac
601 cgcctgatca actgcaacac cagcaccatc acccaggcct gccccaaggt gagcttcgac
661 cccatcccca tccactactg cgcccccgcc ggctacgcca tcctgaagtg caacaacaag
721 accttcaacg gcaccggccc ctgctacaac gtgagcaccg tgcagtgcac ccacggcatc
781 aagcccgtgg tgagcaccca gctgctgctg aacggcagcc tggccgagga gggcatcatc
841 atccgcagcg agaacctgac cgagaacacc aagaccatca tcgtgcacct gaacgagagc
901 gtggagatca actgcacccg ccccaacaac aacacccgca agagcgtgcg catcggcccc
961 ggccaggcct tctacgccac caacgacgtg atcggcaaca tccgccaggc ccactgcaac
1021 atcagcaccg accgctggaa caagaccctg cagcaggtga tgaagaagct gggcgagcac
1081 ttccccaaca agaccatcca gttcaagccc cacgccggcg gcgacctgga gatcaccatg
1141 cacagcttca actgccgcgg cgagttcttc tactgcaaca ccagcaacct gttcaacagc
1201 acctaccaca gcaacaacgg cacctacaag tacaacggca acagcagcag ccccatcacc
1261 ctgcagtgca agatcaagca gatcgtgcgc atgtggcagg gcgtgggcca ggccacctac
1321 gccccccccca tcgccggcaa catcacctgc cgcagcaaca tcaccggcat cctgctgacc
1381 cgcgacggcg gcttcaacac caccaacaac accgagacct tccgccccgg cggcggcgac
1441 atgcgcgaca actggcgcag cgagctgtac aagtacaagg tggtggagat caagcccctg
1501 ggcatcgccc ccaccaaggc caagcgccgc gtggtgcagc gcgagaagcg cgccgtgggc
1561 atcggcgccg tgttcctggg cttcctgggc gccgccggca gcaccatggg cgccgccagc
1621 atcaccctga ccgtgcaggc ccgccagctg ctgagcggca tcgtgcagca gcagagcaac
1681 ctgctgaagg ccatcgaggc ccagcagcac atgctgcagc tgaccgtgtg gggcatcaag
1741 cagctgcagg cccgcgtgct ggccatcgag cgctacctga aggaccagca gctgctgggc
1801 atctggggct gcagcggccg cctgatctgc accaccgccg tgccctggaa cagcagctgg
1861 agcaacaaga gcgagaagga catctgggac aacatgacct ggatgcagtg ggaccgcgag
1921 atcagcaact acaccggcct gatctacaac ctgctggagg acagccagaa ccagcaggag
1981 aagaacgaga aggacctgct ggagctggac aagtggaaca acctgtggaa ctggttcgac
2041 atcagcaact ggccctggta catcaagatc ttcatcatga tcgtgggcgg cctgatcggc
2101 ctgcgcatca tcttcgccgt gctgagcatc gtgaaccgcg tgcgccaggg ctacagcccc
2161 ctgagcttcc agaccctgac ccccagcccc cgcggcctgg accgcctggg cggcatcgag
2221 gaggagggcg gcgagcagga ccgcgaccgc agcatccgcc tggtgagcgg cttcctgagc
2281 ctggcctggg acgacctgcg caacctgtgc ctgttcagct accaccgcct gcgcgacttc
2341 atcctgatcg ccgtgcgcgc cgtggagctg ctgggccaca gcagcctgcg cggcctgcag
2401 cgcggctggg agatcctgaa gtacctgggc agcctggtgc agtactgggg cctggagctg
2461 aagaagagcg ccatcagcct gctggacacc atcgccatca ccgtggccga gggcaccgac
2521 cgcatcatcg agctggtgca gcgcatctgc cgcgccatcc tgaacatccc ccgccgcatc
2581 cgccagggct tcgaggccgc cctgctgtaa ctcgag

FIGURE 98

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccaccacc gagctgcgcg acaagaagca caaggagtac
541 gccctgttct accgcctgga catcgtgccc ctgaacgaga acagcgacaa cttcacctac
601 cgcctgatca actgcaacac cagcaccatc acccaggcct gccccaaggt gagcttcgac
661 cccatcccca tccactactg cgcccccgcc ggctacgcca tcctgaagtg caacaacaag
721 accttcaacg gcaccggccc ctgctacaac gtgagcaccg tgcagtgcac ccacggcatc
781 aagcccgtgg tgagcaccca gctgctgctg aacggcagcc tggccgagga gggcatcatc
841 atccgcagcg agaacctgac cgagaacacc aagaccatca tcgtgcacct gaacgagagc
901 gtggagatca actgcacccg ccccaacaac aacacccgca agagcgtgcg catcggcccc
961 ggccaggcct tctacgccac caacgacgtg atcggcaaca tccgccaggc ccactgcaac
1021 atcagcaccg accgctggaa caagaccctg cagcaggtga tgaagaagct gggcgagcac
1081 ttccccaaca agaccatcca gttcaagccc cacgccggcg gcgacctgga gatcaccatg
1141 cacagcttca actgccgcgg cgagttcttc tactgcaaca ccagcaacct gttcaacagc
1201 acctaccaca gcaacaacgg cacctacaag tacaacggca acagcagcag ccccatcacc
1261 ctgcagtgca agatcaagca gatcgtgcgc atgtggcagg gcgtgggcca ggccacctac
1321 gccccccccca tcgccggcaa catcacctgc cgcagcaaca tcaccggcat cctgctgacc
1381 cgcgacggcg gcttcaacac caccaacaac accgagacct tccgcccccgg cggcggcgac
1441 atgcgcgaca actggcgcag cgagctgtac aagtacaagg tggtggagat caagccccctg
1501 ggcatcgccc ccaccaaggc caagcgccgc gtggtgcagc gcgagaagcg cgccgtgggc
1561 atcggcgccg tgttcctggg cttcctgggc gccgccggca gcaccatggg cgccgccagc
1621 atcaccctga ccgtgcaggc ccgccagctg ctgagcggca tcgtgcagca gcagagcaac
1681 ctgctgaagg ccatcgaggc ccagcagcac atgctgcagc tgaccgtgtg gggcatcaag
1741 cagctgcagg cccgcgtgct ggccatcgag cgctacctga aggaccagca gctgctgggc
1801 atctggggct gcagcggccg cctgatctgc accaccgccg tgccctggaa cagcagctgg
1861 agcaacaaga gcgagaagga catctgggac aacatgacct ggatgcagtg ggaccgcgag
1921 atcagcaact acaccggcct gatctacaac ctgctggagg acagccagaa ccagcaggag
1981 aagaacgaga aggacctgct ggagctggac aagtggaaca acctgtggaa ctggttcgac
2041 atcagcaact ggccctggta catcaagatc ttcatcatga tcgtgggcgg cctgatcggc
2101 ctgcgcatca tcttcgccgt gctgagcatc gtgaaccgcg tgcgccaggg ctacagcccc
2161 ctgagcttcc agaccctgac ccccagcccc cgcggcctgg accgcctggg cggcatcgag
2221 gaggagggcg gcgagcagga ccgcgaccgc agcatccgcc tggtgagcgg cttcctgagc
2281 ctggcctggg acgacctgcg caacctgtgc ctgttcagct accaccgcct gcgcgacttc
2341 atcctgatcg ccgtgcgcgc cgtggagctg ctgggccaca gcagcctgcg cggcctgcag
2401 cgcggctggg agatcctgaa gtacctgggc agcctggtgc agtactgggg cctggagctg
2461 aagaagagcg ccatcagcct gctggacacc atcgccatca ccgtggccga gggcaccgac
2521 cgcatcatcg agctggtgca gcgcatctgc cgcgccatcc tgaacatccc ccgccgcatc
2581 cgccagggct tcgaggccgc cctgctgtaa ctcgag

FIGURE 97

gp160mod.TV1.tpa1 (SEQ ID NO:125)

1 gtcgacgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga
61 gcagtcttcg tttcgcccag cgccagcacc gaggacctgt gggtgaccgt gtactacggc
121 gtgcccgtgt ggcgcgacgc caagaccacc ctgttctgcg ccagcgacgc caaggcctac
181 gagaccgagg tgcacaacgt gtgggccacc cacgcctgcg tgcccaccga ccccaacccc
241 caggagatcg tgctgggcaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggcc
301 gaccagatgc acgaggacgt gatcagcctg tgggaccaga gcctgaagcc ctgcgtgaag
361 ctgaccccc tgtgcgtgac cctgaactgc accgacacca acgtgaccgg caaccgcacc
421 gtgaccggca acagcaccaa caacaccaac ggcaccggca tctacaacat cgaggagatg
481 aagaactgca gcttcaacgc caccaccgag ctgcgcgaca agaagcacaa ggagtacgcc
541 ctgttctacc gcctggacat cgtgcccctg aacgagaaca gcgacaactt cacctaccgc
601 ctgatcaact gcaacaccag caccatcacc caggcctgcc ccaaggtgag cttcgacccc
661 atccccatcc actactgcgc ccccgccggc tacgccatcc tgaagtgcaa caacaagacc
721 ttcaacggca ccggcccctg ctacaacgtg agcaccgtgc agtgcaccca cggcatcaag
781 cccgtggtga gcacccagct gctgctgaac ggcagcctgg ccgaggaggg catcatcatc
841 cgcagcgaga acctgaccga gaacaccaag accatcatcg tgcacctgaa cgagagcgtg
901 gagatcaact gcacccgccc caacaacaac acccgcaaga gcgtgcgcat cggccccggc
961 caggccttct acgccaccaa cgacgtgatc ggcaacatcc gccaggccca ctgcaacatc
1021 agcaccgacc gctggaacaa gaccctgcag caggtgatga agaagctggg cgagcacttc
1081 cccaacaaga ccatccagtt caagccccac gccggcggcg acctggagat caccatgcac
1141 agcttcaact gccgcggcga gttcttctac tgcaacacca gcaacctgtt caacagcacc
1201 taccacagca acaacggcac ctacaagtac aacggcaaca gcagcagccc catcaccctg
1261 cagtgcaaga tcaagcagat cgtgcgcatg tggcagggcg tgggccaggc cacctacgcc
1321 ccccccatcg ccggcaacat cacctgccgc agcaacatca ccggcatcct gctgacccgc
1381 gacggcggct tcaacaccac caacaacacc gagaccttcc gccccggcgg cggcgacatg
1441 cgcgacaact ggcgcagcga gctgtacaag tacaaggtgg tggagatcaa gccctgggc
1501 atcgccccca ccaaggccaa gcgccgcgtg gtgcagcgcg agaagcgcgc cgtgggcatc
1561 ggcgccgtgt tcctgggctt cctgggcgcc gccggcagca ccatgggcgc cgccagcatc
1621 accctgaccg tgcaggcccg ccagctgctg agcggcatcg tgcagcagca gagcaacctg
1681 ctgaaggcca tcgaggccca gcagcacatg ctgcagctga ccgtgtgggg catcaagcag
1741 ctgcaggccc gcgtgctggc catcgagcgc tacctgaagg accagcagct gctgggcatc
1801 tggggctgca gcggccgcct gatctgcacc accgccgtgc cctggaacag cagctggagc
1861 aacaagagcg agaaggacat ctgggacaac atgacctgga tgcagtggga ccgcgagatc
1921 agcaactaca ccggcctgat ctacaacctg ctggaggaca gccagaacca gcaggagaag
1981 aacgagaagg acctgctgga gctggacaag tggaacaacc tgtggaactg gttcgacatc
2041 agcaactggc cctggtacat caagatcttc atcatgatcg tgggcggcct gatcggcctg
2101 cgcatcatct tcgccgtgct gagcatcgtg aaccgcgtgc gccagggcta cagccccctg
2161 agcttccaga ccctgacccc cagccccccgc ggcctggacc gcctgggcgg catcgaggag
2221 gagggcggcg agcaggaccg cgaccgcagc atccgcctgg tgagcggctt cctgagcctg
2281 gcctgggacg acctgcgcaa cctgtgcctg ttcagctacc accgcctgcg cgacttcatc
2341 ctgatcgccg tgcgcgccgt ggagctgctg ggccacagca gcctgcgcgg cctgcagcgc
2401 ggctgggaga tcctgaagta cctgggcagc ctggtgcagt actggggcct ggagctgaag
2461 aagagcgcca tcagcctgct ggacaccatc gccatcaccg tggccgaggg caccgaccgc
2521 atcatcgagc tggtgcagcg catctgccgc gccatcctga acatcccccg ccgcatccgc
2581 cagggcttcg aggccgccct gctgtaactc gag

FIGURE 96

gp160mod.TV1.mut7.delV2 (SEQ ID NO:124)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc
541 atcacccagg cctgccccaa ggtgagcttc gaccccatcc ccatccacta ctgcgccccc
601 gccggctacg ccatcctgaa gtgcaacaac aagaccttca acggcaccgg cccctgctac
661 aacgtgagca ccgtgcagtg cacccacggc atcaagcccg tggtgagcac ccagctgctg
721 ctgaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac
781 accaagacca tcatcgtgca cctgaacgag agcgtggaga tcaactgcac ccgccccaac
841 aacaacaccc gcaagagcgt gcgcatcggc cccggccagg ccttctacgc caccaacgac
901 gtgatcggca acatccgcca ggcccactgc aacatcagca ccgaccgctg gaacaagacc
961 ctgcagcagg tgatgaagaa gctgggcgag cacttcccca acaagaccat ccagttcaag
1021 ccccacgccg gcggcgacct ggagatcacc atgcacagct tcaactgccg cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatcgtg
1201 cgcatgtggc agggcgtggg ccaggccacc tacgcccccc ccatcgccgg caacatcacc
1261 tgccgcagca acatcaccgg catcctgctg acccgcgacg gcggcttcaa caccaccaac
1321 aacaccgaga ccttccgccc cggcggcggc gacatgcgcg acaactggcg cagcgagctg
1381 tacaagtaca aggtggtgga gatcaagccc ctgggcatcg ccccaccaa ggccatcagc
1441 agcgtggtgc agagcgagaa gagcgccgtg ggcatcggcg ccgtgttcct gggcttcctg
1501 ggcgccgccg gcagcaccat gggcgccgcc agcatcaccc tgaccgtgca ggcccgccag
1561 ctgctgagcg gcatcgtgca gcagcagagc aacctgctga aggccatcga ggcccagcag
1621 cacatgctgc agctgaccgt gtggggcatc aagcagctgc aggcccgcgt gctggccatc
1681 gagcgctacc tgaaggacca gcagctgctg ggcatctggg gctgcagcgg ccgcctgatc
1741 tgcaccaccg ccgtgccctg gaacagcagc tggagcaaca agagcgagaa ggacatctgg
1801 gacaacatga cctggatgca gtgggaccgc gagatcagca actacaccgg cctgatctac
1861 aacctgctgg aggacagcca gaaccagcag gagaagaacg agaaggacct gctggagctg
1921 gacaagtgga acaacctgtg gaactggttc gacatcagca actggccctg gtacatcaag
1981 atcttcatca tgatcgtggg cggcctgatc ggcctgcgca tcatcttcgc cgtgctgagc
2041 atcgtgaacc gcgtgcgcca gggctacagc cccctgagct tccagacccc gacccccagc
2101 ccccgcggcc tggaccgcct gggcggcatc gaggaggagg gcggcgagca ggaccgcgac
2161 cgcagcatcc gcctggtgag cggcttcctg agcctggcct gggacgacct gcgcaacctg
2221 tgcctgttca gctaccaccg cctgcgcgac ttcatcctga tcgccgtgcg cgccgtggag
2281 ctgctgggcc acagcagcct gcgcggcctg cagcgcggct gggagatcct gaagtacctg
2341 ggcagcctgg tgcagtactg gggcctggag ctgaagaaga gcgccatcag cctgctggac
2401 accatcgcca tcaccgtggc cgagggcacc gaccgcatca tcgagctggt gcagcgcatc
2461 tgccgcgcca tcctgaacat ccccgccgc atccgccagg gcttcgaggc cgccctgctg
2521 taactcgag

FIGURE 95

gp160mod.TV1.delV2 (SEQ ID NO:123)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc
541 atcacccagg cctgccccaa ggtgagcttc gaccccatcc ccatccacta ctgcgccccc
601 gccggctacg ccatcctgaa gtgcaacaac aagaccttca acggcaccgg cccctgctac
661 aacgtgagca ccgtgcagtg cacccacggc atcaagcccg tggtgagcac ccagctgctg
721 ctgaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac
781 accaagacca tcatcgtgca cctgaacgag agcgtggaga tcaactgcac ccgccccaac
841 aacaacaccc gcaagagcgt gcgcatcggc cccggccagg ccttctacgc caccaacgac
901 gtgatcggca acatccgcca ggcccactgc aacatcagca ccgaccgctg gaacaagacc
961 ctgcagcagg tgatgaagaa gctgggcgag cacttcccca acaagaccat ccagttcaag
1021 ccccacgccg gcggcgacct ggagatcacc atgcacagct tcaactgccg cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatcgtg
1201 cgcatgtggc agggcgtggg ccaggccacc tacgcccccc ccatcgccgg caacatcacc
1261 tgccgcagca acatcaccgg catcctgctg acccgcgacg gcggcttcaa caccaccaac
1321 aacaccgaga ccttccgccc cggcggcggc gacatgcgcg acaactggcg cagcgagctg
1381 tacaagtaca aggtggtgga gatcaagccc ctgggcatcg cccccaccaa ggccaagcgc
1441 cgcgtggtgc agcgcgagaa gcgcgccgtg ggcatcggcg ccgtgttcct gggcttcctg
1501 ggcgccgccg gcagcaccat gggcgccgcc agcatcaccc tgaccgtgca ggcccgccag
1561 ctgctgagcg gcatcgtgca gcagcagagc aacctgctga aggccatcga ggcccagcag
1621 cacatgctgc agctgaccgt gtggggcatc aagcagctgc aggcccgcgt gctggccatc
1681 gagcgctacc tgaaggacca gcagctgctg ggcatctggg gctgcagcgg ccgcctgatc
1741 tgcaccaccg ccgtgccctg gaacagcagc tggagcaaca agagcgagaa ggacatctgg
1801 gacaacatga cctggatgca gtgggaccgc gagatcagca actacaccgg cctgatctac
1861 aacctgctgg aggacagcca gaaccagcag gagaagaacg agaaggacct gctggagctg
1921 gacaagtgga acaacctgtg gaactggttc gacatcagca actggccctg gtacatcaag
1981 atcttcatca tgatcgtggg cggcctgatc ggcctgcgca tcatcttcgc cgtgctgagc
2041 atcgtgaacc gcgtgcgcca gggctacagc ccctgagct tccagaccct gaccccagc
2101 ccccgcggcc tggaccgcct gggcggcatc gaggaggagg gcggcgagca ggaccgcgac
2161 cgcagcatcc gcctggtgag cggcttcctg agcctggcct gggacgacct gcgcaacctg
2221 tgcctgttca gctaccaccg cctgcgcgac ttcatcctga tcgccgtgcg cgccgtggag
2281 ctgctgggcc acagcagcct gcgcggcctg cagcgcggct gggagatcct gaagtacctg
2341 ggcagcctgg tgcagtactg gggcctggag ctgaagaaga gcgccatcag cctgctggac
2401 accatcgcca tcaccgtggc cgagggcacc gaccgcatca tcgagctggt gcagcgcatc
2461 tgccgcgcca tcctgaacat cccccgccgc atccgccagg gcttcgaggc cgccctgctg
2521 taactcgag

FIGURE 94

gp160mod.TV1.delV1V2 (SEQ ID NO:122)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gggcgccggc aactgcaaca ccagcaccat cacccaggcc
421 tgccccaagg tgagcttcga ccccatcccc atccactact gcgccccccgc cggctacgcc
481 atcctgaagt gcaacaacaa gaccttcaac ggcaccggcc cctgctacaa cgtgagcacc
541 gtgcagtgca cccacggcat caagcccgtg gtgagcaccc agctgctgct gaacggcagc
601 ctggccgagg agggcatcat catccgcagc gagaacctga ccgagaacac caagaccatc
661 atcgtgcacc tgaacgagag cgtggagatc aactgcaccc gccccaacaa caacacccgc
721 aagagcgtgc gcatcggccc cggccaggcc ttctacgcca ccaacgacgt gatcggcaac
781 atccgccagg cccactgcaa catcagcacc gaccgctgga acaagaccct gcagcaggtg
841 atgaagaagc tgggcgagca cttccccaac aagaccatcc agttcaagcc ccacgccggc
901 ggcgacctgg agatcaccat gcacagcttc aactgccgcg gcgagttctt ctactgcaac
961 accagcaacc tgttcaacag cacctaccac agcaacaacg gcacctacaa gtacaacggc
1021 aacagcagca gccccatcac cctgcagtgc aagatcaagc agatcgtgcg catgtggcag
1081 ggcgtgggcc aggccaccta cgccccccccc atcgccggca acatcacctg ccgcagcaac
1141 atcaccggca tcctgctgac ccgcgacggc ggcttcaaca ccaccaacaa caccgagacc
1201 ttccgccccg gcggcggcga catgcgcgac aactggcgca gcgagctgta caagtacaag
1261 gtggtggaga tcaagcccct gggcatcgcc cccaccaagg ccaagcgccg cgtggtgcag
1321 cgcgagaagc gcgccgtggg catcggcgcc gtgttcctgg gcttcctggg cgccgccggc
1381 agcaccatgg gcgccgccag catcaccctg accgtgcagg cccgccagct gctgagcggc
1441 atcgtgcagc agcagagcaa cctgctgaag gccatcgagg cccagcagca catgctgcag
1501 ctgaccgtgt ggggcatcaa gcagctgcag gcccgcgtgc tggccatcga gcgctacctg
1561 aaggaccagc agctgctggg catctggggc tgcagcggcc gcctgatctg caccaccgcc
1621 gtgccctgga acagcagctg gagcaacaag agcgagaagg acatctggga caacatgacc
1681 tggatgcagt gggaccgcga gatcagcaac tacaccggcc tgatctacaa cctgctggag
1741 gacagccaga accagcagga gaagaacgag aaggacctgc tggagctgga caagtggaac
1801 aacctgtgga actggttcga catcagcaac tggccctggt acatcaagat cttcatcatg
1861 atcgtgggcg gcctgatcgg cctgcgcatc atcttcgccg tgctgagcat cgtgaaccgc
1921 gtgcgccagg gctacagccc cctgagcttc cagaccctga cccccagccc ccgcggcctg
1981 gaccgcctgg gcggcatcga ggaggagggc ggcgagcagg accgcgaccg cagcatccgc
2041 ctggtgagcg gcttcctgag cctggcctgg gacgacctgc gcaacctgtg cctgttcagc
2101 taccaccgcc tgcgcgactt catcctgatc gccgtgcgcg ccgtggagct gctgggccac
2161 agcagcctgc gcggcctgca gcgcggctgg gagatcctga agtacctggg cagcctggtg
2221 cagtactggg gcctggagct gaagaagagc gccatcagcc tgctggacac catcgccatc
2281 accgtggccg agggcaccga ccgcatcatc gagctggtgc agcgcatctg ccgcgccatc
2341 ctgaacatcc cccgccgcat ccgccagggc ttcgaggccg ccctgctgta actcgag

FIGURE 93

gp140mod.TV1.mut7.delV2 (SEQ ID NO:121)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc
541 atcacccagg cctgccccaa ggtgagcttc gaccccatcc ccatccacta ctgcgccccc
601 gccggctacg ccatcctgaa gtgcaacaac aagaccttca acggcaccgg cccctgctac
661 aacgtgagca ccgtgcagtg cacccacggc atcaagcccg tggtgagcac ccagctgctg
721 ctgaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac
781 accaagacca tcatcgtgca cctgaacgag agcgtggaga tcaactgcac ccgccccaac
841 aacaacaccc gcaagagcgt gcgcatcggc cccggccagg ccttctacgc caccaacgac
901 gtgatcggca acatccgcca ggcccactgc aacatcagca ccgaccgctg gaacaagacc
961 ctgcagcagg tgatgaagaa gctgggcgag cacttcccca acaagaccat ccagttcaag
1021 ccccacgccg gcggcgacct ggagatcacc atgcacagct tcaactgccg cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatcgtg
1201 cgcatgtggc agggcgtggg ccaggccacc tacgcccccc ccatcgccgg caacatcacc
1261 tgccgcagca acatcaccgg catcctgctg acccgcgacg gcggcttcaa caccaccaac
1321 aacaccgaga ccttccgccc cggcggcggc gacatgcgcg acaactggcg cagcgagctg
1381 tacaagtaca aggtggtgga gatcaagccc ctgggcatcg cccccaccaa ggccatcagc
1441 agcgtggtgc agagcgagaa gagcgccgtg ggcatcggcg ccgtgttcct gggcttcctg
1501 ggcgccgccg gcagcaccat gggcgccgcc agcatcaccc tgaccgtgca ggcccgccag
1561 ctgctgagcg gcatcgtgca gcagcagagc aacctgctga aggccatcga ggcccagcag
1621 cacatgctgc agctgaccgt gtggggcatc aagcagctgc aggcccgcgt gctggccatc
1681 gagcgctacc tgaaggacca gcagctgctg ggcatctggg gctgcagcgg ccgcctgatc
1741 tgcaccaccg ccgtgccctg gaacagcagc tggagcaaca agagcgagaa ggacatctgg
1801 gacaacatga cctggatgca gtgggaccgc gagatcagca actacaccgg cctgatctac
1861 aacctgctgg aggacagcca gaaccagcag gagaagaacg agaaggacct gctggagctg
1921 gacaagtgga acaacctgtg gaactggttc gacatcagca actggccctg gtacatctaa
1981 ctcgag

FIGURE 92

gp140mod.TV1.delV2 (SEQ ID NO:120)

```
   1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
  61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
 121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
 181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
 241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
 301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
 361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
 421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
 481 atgaagaact gcagcttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc
 541 atcacccagg cctgccccaa ggtgagcttc gaccccatcc ccatccacta ctgcgccccc
 601 gccggctacg ccatcctgaa gtgcaacaac aagaccttca acggcaccgg cccctgctac
 661 aacgtgagca ccgtgcagtg cacccacggc atcaagcccg tggtgagcac ccagctgctg
 721 ctgaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac
 781 accaagacca tcatcgtgca cctgaacgag agcgtggaga tcaactgcac ccgccccaac
 841 aacaacaccc gcaagagcgt gcgcatcggc cccggccagg ccttctacgc caccaacgac
 901 gtgatcggca acatccgcca ggcccactgc aacatcagca ccgaccgctg gaacaagacc
 961 ctgcagcagg tgatgaagaa gctgggcgag cacttcccca acaagaccat ccagttcaag
1021 ccccacgccg gcggcgacct ggagatcacc atgcacagct tcaactgccg cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatcgtg
1201 cgcatgtggc agggcgtggg ccaggccacc tacgcccccc ccatcgccgg caacatcacc
1261 tgccgcagca acatcaccgg catcctgctg acccgcgacg gcggcttcaa caccaccaac
1321 aacaccgaga ccttccgccc cggcggcggc gacatgcgcg acaactggcg cagcgagctg
1381 tacaagtaca aggtggtgga gatcaagccc ctgggcatcg cccccaccaa ggccaagcgc
1441 cgcgtggtgc agcgcgagaa gcgcgccgtg ggcatcggcg ccgtgttcct gggcttcctg
1501 ggcgccgccg gcagcaccat gggcgccgcc agcatcaccc tgaccgtgca ggcccgccag
1561 ctgctgagcg gcatcgtgca gcagcagagc aacctgctga aggccatcga ggcccagcag
1621 cacatgctgc agctgaccgt gtggggcatc aagcagctgc aggcccgcgt gctggccatc
1681 gagcgctacc tgaaggacca gcagctgctg ggcatctggg gctgcagcgg ccgcctgatc
1741 tgcaccaccg ccgtgccctg gaacagcagc tggagcaaca agagcgagaa ggacatctgg
1801 gacaacatga cctggatgca gtgggaccgc gagatcagca actacaccgg cctgatctac
1861 aacctgctgg aggacagcca gaaccagcag gagaagaacg agaaggacct gctggagctg
1921 gacaagtgga acaacctgtg gaactggttc gacatcagca actggccctg gtacatctaa
1981 ctcgag
```

FIGURE 91

gp120mod.TV1.delV2 (SEQ ID NO:119)

1 gaattcatgc gcgtgatggg cacccagaag aactgccagc agtggtggat ctggggcatc
61 ctgggcttct ggatgctgat gatctgcaac accgaggacc tgtgggtgac cgtgtactac
121 ggcgtgcccg tgtggcgcga cgccaagacc accctgttct gcgccagcga cgccaaggcc
181 tacgagaccg aggtgcacaa cgtgtgggcc acccacgcct gcgtgcccac cgaccccaac
241 ccccaggaga tcgtgctggg caacgtgacc gagaacttca acatgtggaa gaacgacatg
301 gccgaccaga tgcacgagga cgtgatcagc ctgtgggacc agagcctgaa gccctgcgtg
361 aagctgaccc ccctgtgcgt gaccctgaac tgcaccgaca ccaacgtgac cggcaaccgc
421 accgtgaccg gcaacagcac caacaacacc aacggcaccg gcatctacaa catcgaggag
481 atgaagaact gcagcttcaa cgccggcgcc ggccgcctga tcaactgcaa caccagcacc
541 atcacccagg cctgccccaa ggtgagcttc gaccccatcc ccatccacta ctgcgccccc
601 gccggctacg ccatcctgaa gtgcaacaac aagaccttca acggcaccgg ccctgctac
661 aacgtgagca ccgtgcagtg cacccacggc atcaagcccg tggtgagcac ccagctgctg
721 ctgaacggca gcctggccga ggagggcatc atcatccgca gcgagaacct gaccgagaac
781 accaagacca tcatcgtgca cctgaacgag agcgtggaga tcaactgcac ccgccccaac
841 aacaacaccc gcaagagcgt gcgcatcggc cccggccagg ccttctacgc caccaacgac
901 gtgatcggca acatccgcca ggcccactgc aacatcagca ccgaccgctg gaacaagacc
961 ctgcagcagg tgatgaagaa gctgggcgag cacttcccca acaagaccat ccagttcaag
1021 ccccacgccg gcggcgacct ggagatcacc atgcacagct tcaactgccg cggcgagttc
1081 ttctactgca acaccagcaa cctgttcaac agcacctacc acagcaacaa cggcacctac
1141 aagtacaacg gcaacagcag cagccccatc accctgcagt gcaagatcaa gcagatcgtg
1201 cgcatgtggc agggcgtggg ccaggccacc tacgcccccc ccatcgccgg caacatcacc
1261 tgccgcagca acatcaccgg catcctgctg acccgcgacg gcggcttcaa caccaccaac
1321 aacaccgaga ccttccgccc cggcggcggc gacatgcgcg acaactggcg cagcgagctg
1381 tacaagtaca aggtggtgga gatcaagccc ctgggcatcg cccccaccaa ggccaagcgc
1441 cgcgtggtgc agcgcgagaa gcgctaactc gag

FIGURE 90

Vpu_TV2_C_ZAwt (SEQ ID NO:118)

ATGTTAGATTTAACTGCAAGAATAGATTCTAGATTAGGAATAGGAGCATTGA
TAGTAGCACTAATCATAGCAATAATAGTGTGGACCATAGTATATATAGAATA
TAGGAAATTGGTAAGGCAAAGGAAAATAGACTGGTTAGTTAAAAGGATTAG
GGAAAGAGCAGAAGACAGTGGCAATGAGAGCGAGGGGGATACTGAAGAATT
ATCGACACTGGTGGATATGGGGCATCTTAGGCTTTTGGATGCTAATGATGTGT
AA

FIGURE 89

Vpu_TV2_C_ZAopt (SEQ ID NO:117)

ATGCTGGACCTGACCGCCCGCATCGACAGCCGCCTGGGCATCGGCGCCCTGA
TCGTGGCCCTGATCATCGCCATCATCGTGTGGACCATCGTGTACATCGAGTAC
CGCAAGCTGGTGCGCCAGCGCAAGATCGACTGGCTGGTGAAGCGCATCCGCG
AGCGCGCCGAGGACAGCGGCAACGAGAGCGAGGGCGACACCGAGGAGCTGA
GCACCCTGGTGGACATGGGCCACCTGCGCCTGCTGGACGCCAACGACGTGTA
A

FIGURE 88

Vpr_TV2_C_ZAwt (SEQ ID NO:116)

ATGGAACAAGCCCCAGAAGACCAGGGGCCGCAGAGGGAACCATACAATGAA
TGGACACTAGAGCTTTTAGAAGAACTCAAGCAGGAAGCTGTCAGACACTTTC
CTAGACCATGGCTCCATAACTTAGGACAACATATCTATGAAACCTATGGAGA
TACTTGGACAGGAGTTGAAGCAATAATAAGAATCCTGCAACAATTACTGTTT
ATTCATTTCAGGATTGGGTGCCATCATAGCAGAATAGGCATTTTGCGACAGA
GAAGAGCAAGAAATGGAGCCAATAGATCC

FIGURE 87

Vpr_TV2_C_ZAopt (SEQ ID NO:115)

ATGGAGCAGGCCCCCGAGGACCAGGGCCCCCAGCGCGAGCCCTACAACGAG
TGGACCCTGGAGCTGCTGGAGGAGCTGAAGCAGGAGGCCGTGCGCCACTTCC
CCCGCCCCTGGCTGCACAACCTGGGCCAGCACATCTACGAGACCTACGGCGA
CACCTGGACCGGCGTGGAGGCCATCATCCGCATCCTGCAGCAGCTGCTGTTC
ATCCACTTCCGCATCGGCTGCCACCACAGCCGCATCGGCATCCTGCGCCAGC
GCCGCGCCCGCAACGGCGCCAACCGCAGC

FIGURE 86

Vif_TV2_C_ZAwt (SEQ ID NO:114)

ATGGAAAACAGATGGCAGGTGCTGATTGTGTGGCAGGTAGACAGGATGAAG
ATTAGAACATGGCACAGTTTAGTAAAGCACCATATGTATGTTTCGAGGAGAG
CTGATGGATGGTTCTACAGACATCATTATGAAAGCAGACACCCAAAAGTAAG
TTCAGAAGTACACATCCCATTAGGAGATGCCAGGTTAGTAATAAAAACATAT
TGGGGTCTGCAGACAGGAGAAAGAGCTTGGCATTTGGGTCACGGAGTCTCCA
TAGAATGGAGATTGAGAAGATATAGCACACAAGTAGACCCTGACCTGACAG
ACCAACTAATTCATATGCATTATTTTGATTGTTTTGCAGAATCTGCCATAAGG
AAAGCCATACTAGGACAGATAGTTAGCCCTAAGTGTGACTATCAAGCAGGAC
ATAACAAGGTAGGATCTCTACAATACTTGGCACTGACAGCATTGATAAAACC
AAAAAAGATAAAGCCACCTCTGCCTAGTGTTAGGAAATTAGTAGAGGATAGA
TGGAACAAGCCCCAGAAGACCAGGGGCCGCAGAGGGAACCATACAATGAAT
GGACACTAG

FIGURE 85

Vif_TV2_C_ZAopt (SEQ ID NO:113)

ATGGAGAACCGCTGGCAGGTGCTGATCGTGTGGCAGGTGGACCGCATGAAGA
TCCGCACCTGGCACAGCCTGGTGAAGCACCACATGTACGTGAGCCGCCGCGC
CGACGGCTGGTTCTACCGCCACCACTACGAGAGCCGCCACCCCAAGGTGAGC
AGCGAGGTGCACATCCCCCTGGGCGACGCCCGCCTGGTGATCAAGACCTACT
GGGGCCTGCAGACCGGCGAGCGCGCCTGGCACCTGGGCCACGGCGTGAGCA
TCGAGTGGCGCCTGCGCCGCTACAGCACCCAGGTGGACCCCGACCTGACCGA
CCAGCTGATCCACATGCACTACTTCGACTGCTTCGCCGAGAGCGCCATCCGC
AAGGCCATCCTGGGCCAGATCGTGAGCCCCAAGTGCGACTACCAGGCCGGCC
ACAACAAGGTGGGCAGCCTGCAGTACCTGGCCCTGACCGCCCTGATCAAGCC
CAAGAAGATCAAGCCCCCCCTGCCCAGCGTGCGCAAGCTGGTGGAGGACCGC
TGGAACAAGCCCCAGAAGACCCGCGGCCGCCGCGGCAACCACACCATGAAC
GGCCACTAG

FIGURE 84

TatExon2_TV2_C_ZAwt (SEQ ID NO:112)

CCCTTATCCCAAACCCGAGGGGACCCGACAGGCTCGGAGGAATCGAAGAAG
AAGGTGGAGAGCAAGACAGCAGCAGATCCATTCGATTAG

FIGURE 83

TatExon2_TV2_C_ZAopt (SEQ ID NO:111)

CCCCTGAGCCAGACCCGCGGCGACCCCACCGGCAGCGAGGAGAGCAAGAAG
AAGGTGGAGAGCAAGACCGCCGCCGACCCCTTCGACTAG

FIGURE 82

TatExon1_TV2_C_ZAwt (SEQ ID NO:110)

ATGGAGCCAATAGATCCTAACCTAGAACCCTGGAACCATCCAGGAAGTCAGC
CTAAAACTGCTTGTAATGGGTGTTACTGTAAACGTTGCAGCTATCATTGTCTA
GTTTGCTTTCAGAAAAAAGGCTTAGGCATTTACTATGGCAGGAAGAAGCGGA
GACAGCGACGAAGCGCTCCTCCAAGCAATAAAGATCATCAAGATCCTCTACC
AAAGCAG

FIGURE 81

TatExon1_TV2_C_ZAopt (SEQ ID NO:109)

ATGGAGCCCATCGACCCCAACCTGGAGCCCTGGAACCACCCCGGCAGCCAGC
CCAAGACCGCCTGCAACGGCTGCTACTGCAAGCGCTGCAGCTACCACTGCCT
GGTGTGCTTCCAGAAGAAGGGCCTGGGCATCTACTACGGCCGCAAGAAGCGC
CGCCAGCGCCGCAGCGCCCCCCCCAGCAACAAGGACCACCAGGACCCCCTGC
CCAAGCAG

FIGURE 80

RevExon2_TV2_C_ZAwt (SEQ ID NO:108)

ACCCTTATCCCAAACCCGAGGGGACCCGACAGGCTCGGAGGAATCGAAGAA
GAAGGTGGAGAGCAAGACAGCAGCAGATCCATTCGATTAGTGAGCGGATTCT
TGACACTTGCCTGGGACGACCTACGAAGCCTGTGCCTCTTCTGCTACCACCGA
TTGAGAGACTTCATATTAATTGTAGTGAGAGCAGTGGAACTTCTGGGACACA
GTAGTCTCAGGGGACTGCAGAGGGGGTGGGGAACCCTTAA

FIGURE 79

RevExon2_TV2_C_ZAopt (SEQ ID NO:107)

CCCTACCCCAAGCCCGAGGGCACCCGCCAGGCCCGCCGCAACCGCCGCCGCC
GCTGGCGCGCCCGCCAGCAGCAGATCCACAGCATCAGCGAGCGCATCCTGGA
CACCTGCCTGGGCCGCCCCACCAAGCCCGTGCCCCTGCTGCTGCCCCCCATCG
AGCGCCTGCACATCAACTGCAGCGAGAGCAGCGGCACCAGCGGCACCCAGT
AGAGCCAGGGCACCGCCGAGGGCGTGGGCAACCCCTAA

FIGURE 78